DTIC FILE COPY

AD-A197 432

CHANGE POINT PROBLEMS IN REGRESSION

by

Hyune-Ju Kim
Stanford University

TECHNICAL REPORT NO. 7
JUNE 1988

PREPARED UNDER CONTRACT
N00014-87-K-0078 (NR-042-373)
FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

CHANGE POINT PROBLEMS IN REGRESSION

by

Hyune–Ju Kim
Stanford University

TECHNICAL REPORT NO. 7
JUNE 1988

PREPARED UNDER CONTRACT
N00014-87-K-0078 (NR-042-373)
FOR THE OFFICE OF NAVAL RESEARCH

Also prepared under National Science Foundation grant DMS86-002335 and issued as Technical Report No. 285, Department of Statistics, Stanford University.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

CHANGE-POINT PROBLEMS IN REGRESSION

Hyune-Ju Kim, Ph.D.

Stanford University, 1988

This dissertation focuses on the problem of testing for a change in the regression model when errors are independently, normally distributed with constant, known or unknown variance. First we consider the regression model in which only the intercept changes at some unknown point (Model-1). Secondly, the model in which both intercept and slope change is considered (Model-2). In all cases, the likelihood ratio statistic (LRS) is of the form $U = \max_{1 \leq i < m} U_i$, where distributions of U_i's vary according to the assumptions.

In both models, ~~we~~ consider the likelihood ratio test (LRT) as the problem of the boundary crossing by the discrete stochastic process and ~~study~~ problems such as approximations to significance levels, powers, and confidence regions for a change point. First of all, ~~we~~ propose a modified LRT and discuss asymptotic properties of test statistics in cases of random and fixed independent variables. In both cases, ~~we~~ derive analytical approximations to significance levels. When the independent variables are random, the limiting distribution of the modified LRS is a function of a Brownian motion and approximations in Siegmund (1986, Annals of Statistics) are used. For fixed independent variables, the limiting distribution involves a Gaussian process with nondifferentiable sample paths. In this case, an approximation is derived assuming the known variance and mild conditions about the empirical distribution of the independent variable, using the argument in Leadbetter, Lindgren and Rootzen (1983,Chapter12), modified for discrete time by Hogan and Siegmund (1986, Advances in Applied Mathematics). In Model-1, we are also concerned with the power of the LRT and confidence regions for a change point.

Numerical approximations of significance levels and powers of the LRT and the results of corresponding Monte Carlo experiments are obtained. We find that the simulations confirm that the theoretical results perform well and demonstrate that the results also can be applied to the unknown variance case.

Acknowledgments

I would like to thank Professor David Siegmund for being a great dissertation advisor. He suggested this research topic and provided me with valuable ideas and insights, unending patience and encouragement, and careful reading of my work throughout the writing of this dissertation. It has been a great pleasure to be a student of Professor Siegmund whom I will remember as an enthusiastic researcher, wonderful teacher, and sincere advisor with my Stanford life.

I also thank Professors T.L.Lai and Iain Johnstone for reading the manuscript and many helpful comments.

I am particularly grateful to my parents, Donghyuck Kim and Kwangsoon Lee, and my dear husband Youngbai Moon for their love and support.

Finally, I thank my Lord for giving me the opportunity to study at Stanford and for guiding me all through my life.

Table of Contents

chapter 1

Introduction

1.1. Change-Point Problems

In recent years increasing interest has been shown in problems about stability of models for a sequence of observations. When a series of observations is taken sequentially, it can happen that the whole set of observations can be divided into subsets, each of which can be regarded as a random sample from a different distribution. Assuming points at which model changes are unknown, basically two distinct problems arise : detection and estimation of change points.

Change-point problems originally arose in quality control to detect changes in the quality of output from a continuous production process. A process in control maintains an approximately constant quality of output. Suppose that the process jumps out of control at some unknown point, the quality worsens and the output become unacceptable. In order to take actions when such a deterioration is suspected, it is required to signal any departure of the output from the target value as soon as possible.

One of the simplest examples is the problem of detecting a single change in the mean of normal random variables having known and fixed variance. Sequential detection of a change in the mean of the distribution of observations has been studied by Page (1954), Shiryayev (1963), Lorden (1971), and Pollak (1985). For fixed sample problems involving a finite sequence of observations, Siegmund (1986) gave an analytic approximation for a significance level of the likelihood ratio test (LRT) and discussed confidence sets for a change point. James, James, and Siegmund (1987) considered the unknown variance case as well as the known variance case and studied various tests, such as those based on the likelihood ratio and recursive residuals. Also power approximations were developed by integrating approximations for conditional boundary crossing probabilities.

Change-point problems arise in various ways and have been considered in regression models, time-series models, and survival analysis. For a change in a binomial probability, Hinkley and Hinkley (1970) used maximum likelihood methods to estimate a change point for binary random variables and derived exact and asymptotic distributions of the maximum likelihood estimator of the change point. A cumulative sum test statistic for this problem was proposed by Pettitt (1980) and a nonparametric cumulative sum statistic was applied to binomial random variables by Pettitt (1979). An example of this type of a change in epidemiology was described in Worsley (1983), who used the LRT to test for a change in probability of a sequence of independent binomial random variables. He also compared powers of the LRT and the cumulative sum test and discussed the relationship between the cumulative sum test and a two-sample Kolmogorov-Smirnov test. Worsley (1986) used maximum likelihood methods to test for a change in a sequence of independent random variables from an exponential family. He found the exact null and alternative distributions of the test statistics using an iterative numerical procedure. Exact and approximate confidence regions for the change point were given, based on a level α LRT and a modification of the method proposed by Cox and Spijϕtvoll (1982). He also discussed an application to the data set on the time intervals between explosions in British coal mines between 1875 and 1950.

Change-point problems in time-series models have been considered in Picard (1985) who discussed applications to Canadian lynx data, IBM common stock closing prices, and German unemployment data. Picard was concerned with detecting two kinds of changes: first is a change in the spectrum of a time series; secondly she considered a change in the mean or covariance of an autoregressive process. Matthews, Farewell, and Pyke (1985) gave an example of change-point problems in survival analysis. They considered the problem of testing for a constant failure rate against alternatives with failure rates involving a single change-point. Examples of change-point problems in regression models were discussed in a number of papers. With three econometric examples, Brown, Durbin, and Evans (1975) discussed two-phase multiple regression problems. In Esterby and Elshaarawi (1981), a two-phase polynomial regression model has been proposed for the pollen concentration in lake sediment cores. Also Beals (1972, Chapter 12) shows a data set to which a multi-phase regression model can be applied. In this dissertation, we study change-point problems in regression models, especially two-phase linear regression problems.

1.2. Two-Phase Regression

Regression models which are composed of two different linear phases have many applications. As in Brown, Durbin and Evans (1975), it might be suspected that the slope and the intercept have changed after an unknown point in the sequence of observations. In some cases, it may be necessary to consider a regression model in which only one of the parameters changes, while the other remains constant. Maronna and Yohai (1978) considered a two-phase regression model in which only the intercept term changes and discussed applications in meteorology.

The two-phase regression model was first studied by Quandt (1958) who proposed a maximum likelihood method to estimate the parameters in the broken line regression model. Quandt (1960) also proposed a likelihood ratio test (LRT) to test for a change in the regression model as opposed to the null hypothesis that the data follow only one simple linear regression. On the basis of the empirical distribution resulting from some sampling experiments, he concluded that -2log(likelihood ratio) could not have a chi-square distribution with the appropriate degrees of freedom under the null hypothesis.

A second approach to the problem of testing for a change in a regression model is to use recursive residuals introduced by Brown, Durbin, and Evans (1975). Brown, Durbin, and Evans developed tests based on the cusum and cusum of squares of recursive residuals, defined to be uncorrelated with zero means and constant variance. They also considered other techniques based on moving regressions and on the regression models whose coefficients are polynomial in time. As well, the plotting of Quandt's log likelihood ratio statistic (LRS) was suggested. They discussed applications of these techniques to three sets of real data taken from the field of economics.

Since the 1960's, there has been considerable attention to the estimation of parameters as well as the problem of testing for a change in the regression model. Feder (1975) showed by example that if the true model contains fewer phases than the assumed model, the least squares estimators are not asymptotically normal and the -2log(likelihood ratio) statistic is not asymptotically chi-square. He also concluded that the asymptotic null distribution of the -2log(likelihood ratio) would depend on the configuration of the values of the independent variable. Beckman and Cook (1979) further investigated the dependence of the test on the values of the independent variable and gave critical values for testing for a change in the regression model by simulation. They used 4-different configurations of the values of the independent variable, and their results show that this configuration can have a significant influence on the null distribution of the LRS. They also discussed differences between the continuous model in which the composite regression function is constrained to be continuous at the change point and the discontinuous model in which it is not. Hawkins (1980) pointed out that the inferential theory of the two-phase regression model depends strongly on whether or not continuity at the change-point is assumed.

Difficulties of this problem are the facts that standard maximum likelihood asymptotic theory is not applicable and also the null distribution of the test statistic depends on the spacings of the values of the independent variable. The sampling distributions of most of the test statistics described below are quite complicated. Because of this complexity, most previous work has used numerical or Monte Carlo methods. In 1983, Worsley gave analytic approximations to an upper bound on the null distribution function of the test statistic based on an improved Bonferroni's inequality. He considered a general multiple regression model with a normal random error of constant variance, where there may be a change in the coefficient vector at an unknown point in the data. Worsley's upper bounds are much better than Bonferroni's. However it requires considerable numerical work and sometimes the errors are quite substantial , especially for larger sample sizes.

This dissertation focuses on the problem of testing for a change in the regression model when the errors are independently, normally distributed with constant variance. In this dissertation, two kinds of models are considered. First is the regression model in which only the intercept changes at some unknown point (Model-1). Secondly, the model in which both the intercept and the slope change is considered (Model-2). Model-2 is considered without continuity constraint. The nature of the null distributions of these cases are as follows : In Model-1, if the variance is known, then the LRS is the maximum absolute value of correlated standard normal random variables. If the variance is unknown, then the LRS is the maximum absolute value of the ratios of correlated standard normal random variables and the square root of a chi-square random variable. In Model-2, if the variances of the error variable is known, then the LRS is the maximum of correlated chi-square random variables with 2 degrees of freedom. If the variance is unknown, then the LRS is the maximum of correlated Beta random variables. In all cases, the LRS is of the form

$$U = \max_{1 \leq i < m} U_i,$$

where distributions of U_i's vary according to the assumptions. A point of interest is how to deal with the maximization in the LRS. Since it is difficult to get the exact distribution of U, Beckman and Cook (1979) suggested a simple bound on the distribution function based on Bonferroni's inequality :

$$\Pr(U > u) = \Pr(\bigcup_i A_i) \leq \sum_i \Pr(A_i),$$

where A_i is the event that $\{U_i > u\}$. Worsley (1983) improved this upper bound by

$$\Pr(U > u) = \Pr(\bigcup_i A_i) \leq \sum_i \Pr(A_i) - \sum_i \Pr(A_i \bigcap A_{i+1}),$$

In this dissertation, the LRT is considered as the problem of the boundary crossing by the discrete stochastic process and an approximation to the null distribution function is derived under mild conditions.

Chapter 2 deals with the case that only the intercep· ᵔan change and is organized as follows. In Section 2.1, the modified LRT (MLRT) to test for a change only in the intercept term is proposed. Section 2.2 discusses asymptotic properties of test statistics in the cases of random and fixed independent variables. In both cases, Section 2.3 gives analytic approximations to significance levels. When the independent variable is random, the limiting distribution of the modified LRS (MLRS) involves a Brownian motion and results in Siegmund (1986) are used to approximate significance levels. For fixed values of the independent variable, an approximation is derived assuming that the variance of the error variable is known and that the observations of the independent variable satisfy certain conditions. Since the independent variables are not random in most applications, this case is the most important and the most difficult one. When the independent variables are nonrandom, the limiting distribution of the MLRS is not a function of a well-known process like Brownian motion. However it involves a Gaussian process with nondifferentiable sample paths. To approximate the boundary crossing probability by a discrete stochastic process whose limiting process has a non-differentiable sample path, the argument in Leadbetter, Lindgren and Rootzen (1983,Chapter12), modified for discrete time by Hogan and Siegmund (1986), is used. Section 2.4 is concerned with power of the MLRT and confidence regions for a change point.

Chapter 3 obtains results like those of Chapter 2 for the case in which both the intercept and the slope change.

In Chapters 2 and 3, numerical approximations of significance levels and powers of the MLRT and the results of corresponding Monte Carlo experiments are also reported. The simulations confirm that the theoretical results perform well and demonstrate that the results derived under the assumption that variance is known also can be applied to the unknown variance case.

Finally, the Appendix reviews several basic facts concerning the convergence of stochastic processes and discusses Siegmund's (1986) results which are used in Chapters 2 and 3.

Chapter 2

Change in Intercept Alone

2.1. Models and Test Statistics

Let (x_j, y_j), $j = 1, \ldots, m$, be a sequence of m pairs of observations such that $y_j = \alpha^{(j)} + \beta x_j + \varepsilon_j$, where $\alpha^{(j)}$'s and β are unknown parameters and ε_j's are independently and normally distributed with mean 0 and constant variance σ^2.

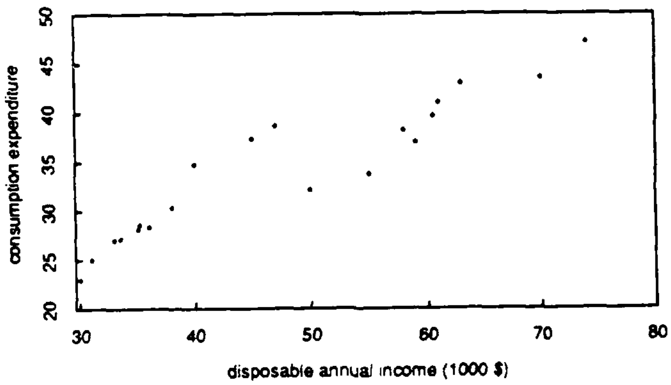
Consider the problem of testing the null hypothesis that the data follow one simple linear regression against the alternative hypothesis that there is a change only in the intercept term. Then the hypotheses can be described more formally as

$$
\begin{aligned}
H_0 &: \alpha^{(j)} = \alpha, \qquad j = 1, \ldots, m, \\
H_1 &: \exists \; 1 \le \rho < m \quad \text{such that} \\
&\quad \alpha^{(j)} = \alpha_1, \qquad j = 1, \ldots, \rho, \\
&\quad \alpha^{(j)} = \alpha_2, \qquad j = \rho + 1, \ldots, m,
\end{aligned}
$$

where $\alpha_1 \neq \alpha_2$.

For the simple case of $\beta = 0$, this problem becomes a test for a single change in the mean of normal random variables with constant variance. Many papers have investigated this type of change-point problem, in particular Gardner (1969), Hinkley (1970), Hawkins (1977), Siegmund (1986), and James, James, and Siegmund (1987). Now if there is a covariate which has a constant effect on the y_j's, the two-phase regression model introduced above could describe the situation. This kind of two-phase regression model can be used to describe the relationship between household consumption and disposable income by the household. Household consumption cannot be explained simply by disposable income of

Fig.1.



the household. Many other variables such as age, sex, race, and education of the family head, may also affect the level of consumption expenditures of the household. For example, consumption patterns according to the age of the family head may be very different. If a sample survey of a household was made for a period including the year when the children of the family began to live independently, then the data might be plotted as in Fig.1.

This testing problem was first studied by Maronna and Yohai (1978). They studied the LRT and also discussed some applications in meteorology. In the next section their approach and some results will be discussed.

In this section, we derive the LRS for cases of known and unknown σ^2 and study the null and alternative distributions of the LRS. When σ^2 is known, σ^2 can be assumed to be equal to 1 without loss of generality. Then -2log(likelihood ratio) statistic for testing H_0 against the alternative that a change occurred at i is proportional to

$$| U_m(i/m) | = \left\{ \frac{mi}{m-i} \right\}^{\frac{1}{2}} \frac{| Q_{xx,m}(\bar{y}_m - \bar{y}_i) - Q_{xy,m}(\bar{x}_m - \bar{x}_i) |}{\{Q^2_{xx,m} - Q_{xx,m}(\bar{x}_m - \bar{x}_i)^2 mi/(m-i)\}^{\frac{1}{2}}}$$

$$= | \hat{\alpha}_i - \hat{\alpha}^*_i | \{[1 - (\bar{x}_m - \bar{x}_i)^2 mi/\{(m-i)Q_{xx,m}\}]m/\{i(m-i)\}\}^{-\frac{1}{2}}.$$

where

$$\bar{x}_i = (\sum_{j=1}^{i} x_j)/i, \qquad \bar{y}_i = (\sum_{j=1}^{i} y_j)/i,$$

$$\bar{x}_i^* = (\sum_{j=i+1}^{m} x_j)/(m-i), \qquad \bar{y}_i^* = (\sum_{j=i+1}^{m} y_j)/(m-i),$$

$$Q_{xx,m} = \sum_{j=1}^{m} (x_j - \bar{x}_m)^2, \qquad Q_{xy,m} = \sum_{j=1}^{m} (x_j - \bar{x}_m)(y_j - \bar{y}_m),$$

$$\hat{\beta} = Q_{xy,m}/Q_{xx,m}, \qquad \hat{\alpha}_i = \bar{y}_i - \hat{\beta}\bar{x}_i, \qquad \hat{\alpha}_i^* = \bar{y}_i^* - \hat{\beta}\bar{x}_i^*.$$

Hence the likelihood ratio test (LRT) of H_0 against H_1 can be based on

$$\max_{1 \leq i < m} \mid U_m(i/m) \mid .$$

Slightly more generally, we shall consider the test statistic

$$M_1 = \max_{m_0 \leq i \leq m_1} \mid U_m(i/m) \mid , \tag{2.1}$$

where $1 < m_0 < m_1 < m$. We will call M_1 as a modified likelihood ratio statistic (MLRS) and the test based on M_1 as a modified likelihood test (MLRT). The MLRT was introduced by Siegmund (1986) who used the MLRS to test for a change in the mean of a sequence of normal random variables. The introduction of m_0 and m_1 in (2.1) can be justified in terms of the power of the test. Since it is intrinsically difficult to detect a change occurring near either of the two end points, the LRS pays for its efforts to detect such a change by having less power at other points. This will be more completely discussed in Section 2.4 with numerical results. Based on the MLRS, H_0 is rejected when M_1 is larger than some constant. The value of i which maximizes $\mid U_m(i/m) \mid$ is the maximum likelihood estimate of the true change point.

Even though the assumption of normal random errors with known variance simplifies this problem, theoretical properties of M_1 are still difficult to characterize. Under H_0, $\hat{\alpha}_i - \hat{\alpha}_i^*$ has a normal distribution with mean 0 and variance $[1 - (\bar{x}_m - \bar{x}_i)^2 mi/\{(m - i)Q_{xx,m}\}]m/\{i(m-i)\}$, and so $U_m(i/m)$ has a standard normal distribution for each i. Hence the null distribution of M_1 is the maximum absolute value of a sequence of correlated standard normal random variables. The covariance between $U_m(i/m)$ and $U_m(j/m)$ for $i < j$ is given by

$$\text{Cov}\,[\ U_m(i/m), U_m(j/m)\ ] = \left\{\frac{(i/m)(1-i/m)}{(j/m)(1-j/m)}\right\}^{\frac{1}{2}} \frac{D_m(i/m, j/m)}{\{D_m(i/m, i/m)D_m(j/m, j/m)\}^{\frac{1}{2}}}, \tag{2.2}$$

where

$$D_m(i/m, j/m) = 1 - (\bar{x}_m - \bar{x}_i)(\bar{x}_m - \bar{x}_j)mj/[(m-j)Q_{xx,m}] \qquad \text{for } i < j.$$

The derivation of (2.2) will be given in the following section. The null distribution of M_1 depends on the x_j's only through this covariance structure of $\{U_m(i/m)\}$, not on α, β. Under the alternative, $U_m(i/m)$ is normally distributed and $\text{Cov}[\ U_m(i/m), U_m(j/m)\ ]$ remains same as under the null. But now $U_m(i/m)$ has non-zero mean for all i, which is given by

$$\begin{aligned} E[U_m(i/m)] &= \frac{i(1-\rho/m)D_m(i/m, \rho/m)}{\{i(1-i/m)D_m(i/m, i/m)\}^{\frac{1}{2}}}\,(\alpha_2 - \alpha_1), \qquad i \le \rho \\ &= \frac{\rho(1-i/m)D_m(i/m, \rho/m)}{\{i(1-i/m)D_m(i/m, i/m)\}^{\frac{1}{2}}}\,(\alpha_2 - \alpha_1), \qquad i > \rho. \end{aligned} \tag{2.3}$$

So the alternative distribution of M_1 depends on the unknown parameter $\alpha_2 - \alpha_1$ and the unknown change point ρ. One interesting property of the test statistic is that a nuisance parameter ρ is present only under the alternative. This property makes analysis difficult since the standard chi-square approximation of -2log(likelihood ratio) can not be applied in this case.

If σ^2 is unknown, the LRS is proportional to

$$\max_{1 \le i < m} \left\{\frac{m^2 i}{m-i}\right\}^{\frac{1}{2}} \frac{|\ Q_{xx,m}(\bar{y}_m - \bar{y}_i) - Q_{xy,m}(\bar{x}_m - \bar{x}_i)\ |}{\{[Q^2_{xx,m} - Q_{xx,m}(\bar{x}_m - \bar{x}_i)^2 mi/(m-i)](Q_{yy,m} - Q^2_{xy,m}/Q_{xx,m})\}^{\frac{1}{2}}}$$

$$= \max_{1 \le i < m} |U_m(i/m)|/\hat{\sigma}\ ,$$

where $\hat{\sigma} = (Q_{yy,m} - Q^2_{xy,m}/Q_{xx,m})/m$.

Again we consider the generalization

$$M_2 = \max_{m_0 \leq i \leq m_1} \mid U_m(i/m) \mid / \hat{\sigma} = \max_{m_0 \leq i \leq m_1} \mid \tilde{U}_m(i/m) \mid,$$

where $\tilde{U}_m(i/m) = U_m(i/m) / \hat{\sigma}$

Under H_0, $U_m(i/m)$ has a standard normal distribution and $\hat{\sigma}^2 m$ has a chi-square distribution with m-2 degrees of freedom. Since $U_m(i/m)$ and $\hat{\sigma}$ are not independent in general and the distribution of $U_m(i/m)$ depends on the x_j's through the complicated covariance structure, it is difficult to find the exact distribution of M_2.

The dependence of test statistics on the values of the independent variable is one of difficulties that must be handled as well as the maximization involved in the definition of the test statistics. By simulation, Beckman and Cook (1979) pointed out that the influence of the configuration of the values of the independent variable is non-negligible and the percentiles of the test statistic increase as the variance of the configuration increases. In the following sections, we will study the asymptotic behavior of the MLRS, especially behavior of the significance level, and will discuss the effect of the spacings of the values of the independent variable.

2.2. Asymptotic Behavior of Test Statistics

In this section, we study asymptotic properties of test statistics when the independent variable is random as well as fixed. The regression model which involves a random independent variable was introduced by Maronna and Yohai (1978). This model is appropriate when the dependent variable may undergo a systematic change at some unknown point, while the independent variable does not change and affects the dependent variable through the correlation between the independent and dependent variable. Maronna and Yohai gave an example of such a situation in meteorology, as follows. Let x and y be two nearby meteorological stations. The measurements might be mean annual precipitations and it might be desired to test the hypothesis that the only fluctuations are those due to the intrinsic randomness of the magnitude being measured, against the alternative that a systematic change has occurred at one of the stations after some point, due to unregistered changes in the measurement apparatus or the location of the station.

In Section 2.2.1, we study the case in which the independent variable is random. Also, the asymptotic behavior of the MLRS is considered conditionally on the x_j's. Section 2.2.2 deals with the case of fixed values of the independent variable. Starting from the special case where the values of the independent variable are uniformly spaced, as they would be if the independent variable is time and observations are made at equal intervals of time, we study the limiting behavior of the MLRS under a mild assumption about the empirical distribution of the independent variable. In Section 2.1, the LRS was derived assuming that the ϵ_j's are identically and normally distributed. The asymptotic results to be discussed in Sections 2.2 and 2.3 holds even in the case of a general error distribution.

2.2.1. When the independent variable is random

Maronna and Yohai (1978) considered the case in which both the independent and dependent variables are random and they studied the limiting distribution under the null hypothesis. Since the LRS does not depend on the slope under the null hypothesis, the independent variable can be taken to be independent of the dependent variable. They gave the percentiles of the LRS when (x, y) has a bivariate normal distribution with $\mathbf{0}$ mean vector and identity covariance matrix, obtained by the Monte Carlo method. Their main result is about the limiting distribution of the test statistic, which will be stated in the following theorem. It was shown that the LRS tends to ∞ as $m \to \infty$ in their paper. Here, we consider the MLRS and show the convergence of the MLRS in distribution. Basically this theorem was proved by Maronna and Yohai, but their proof is not complete in some of the details concerned with the convergence of the stochastic process. In our proof, we consider the "convergence in distribution" in the space $\mathbf{C} = \mathbf{C}[0, 1]$ of continuous functions on $[0, 1]$, equipped with a σ-field $\mathcal{C}$ and the uniform metric.

Notation. Let $W(i/m)$ be a discrete time stochastic process defined at $i = 1, \ldots, m$. Then W^c denotes a process which is continuous in $[0, 1]$, equals W at i/m $(i = 1, \ldots, m)$ and is linear in each interval $(i/m, (i+1)/m)$.

Lemma 2.2.1.

Let $\{v_j\}$ be a sequence of i.i.d. random variables such that $E[{v_j}^2] = 1$. Define $W_m^0(i/m) = (\bar{v}_i - \bar{v}_m)i/\sqrt{m}$, where $\bar{v}_i = (\sum_{j=1}^{i} v_j)/i$. Then as $m \to \infty$,

$$W_m^{0,c} \to W^0 \qquad \text{in distribution,} \tag{2.4}$$

where W^0 is a Brownian bridge process.

Proof : Define $W_m(i/m) = \bar{v}_i i/\sqrt{m}$ and $W_m^c(t)$ to be a continuous process constructed by linear interpolation. By Donsker's Theorem, $W_m^c \to W$ in distribution, where W is a standard Brownian motion. Since a mapping H such that $H(W_m^c) = W_m^{0,c}$ and $H(W) = W^0$ is continuous, by the continuous mapping theorem of weak convergence, (2.4) holds. ∎

Theorem 2.2.2.

Let $(x_1, y_1), \ldots, (x_m, y_m)$ be i.i.d. random variables such that $E[{x_1}^2] < \infty$ and $E[{y_1}^2] < \infty$.

Under H_0, as $m \to \infty$ and $i/m \to t$,

$$U_m(i/m) = \frac{U_m^0(i/m)}{\{(i/m)(1 - i/m)\}^{\frac{1}{2}}} \to \frac{W^0(t)}{\{t(1-t)\}^{\frac{1}{2}}} \qquad \text{in distribution,}$$

where W^0 is a Brownian bridge process.

And so, as $m \to \infty$ and $m_i/m \to t_i$ for $i = 0, 1$,

$$M_1 = \max_{m_0 \le i \le m_1} \frac{|\, U_m^0(i/m)\, |}{\{(i/m)(1 - i/m)\}^{\frac{1}{2}}} \to \max_{t_0 \le t \le t_1} \frac{|\, W^0(t)\, |}{\{t(1-t)\}^{\frac{1}{2}}} \quad \text{in distribution.} \tag{2.5}$$

Proof :

(i) Note that $U_m^0(i/m)$ can be rewritten as

$$[B_y(i/m) - B_x(i/m)Q_{xy,m}/Q_{xx,m}]/\{D_m(i/m, i/m)\}^{\frac{1}{2}},$$

where

$$B_y(i/m) = (\bar{y}_m - \bar{y}_i)i/\sqrt{m},$$

$$B_x(i/m) = (\bar{x}_m - \bar{x}_i)i/\sqrt{m},$$

$$D_m(i/m, i/m) = 1 - [B_x(i/m)]^2/\{Q_{xx,m}(i/m)(1 - i/m)\}.$$

(ii) It may be assumed without loss of generality that $E[x_i] = E[y_i] = 0$ and $E[x_i^2] = E[y_i^2] = 1$ and $\beta = 0$.

Then the Law of Large Numbers implies that

$$Q_{xx,m}/m \to 1, \quad Q_{xy,m}/m \to 0 \qquad \text{in probability.}$$

By the previous lemma,

$$B_x^c \to W_1^0, \quad B_y^c \to W_2^0 \qquad \text{in distribution,}$$

and hence

$$D_m \to 1 \qquad \text{in probability,}$$

where W_1^0 and W_2^0 are two independent Brownian bridges.

Then the continuous mapping theorem implies that as $m \to \infty$,

$$U_m^{0,c} \to W^0 \qquad \text{in distribition.}$$

(iii) Using the continuity of the mapping from Z to $\max_{t_0 \le t \le t_1} |Z(t)|/\{t(1-t)\}^{\frac{1}{2}}$ for $Z \in \mathcal{C}$,

$$\max_{t_0 \le t \le t_1} \frac{|U_m^{0,c}(t)|}{\{t(1-t)\}^{\frac{1}{2}}} \to \max_{t_0 \le t \le t_1} \frac{|W^0(t)|}{\{t(1-t)\}^{\frac{1}{2}}} \qquad \text{in distribution, as } m \to \infty.$$

(iv) However, since

$$\max_{t_0 \le t \le t_1} \frac{|U_m^{0,c}(t)|}{\{t(1-t)\}^{\frac{1}{2}}} \neq \max_{m_0 \le i \le m_1} \frac{|U_m^0(i/m)|}{\{(i/m)(1-i/m)\}^{\frac{1}{2}}},$$

we still need to show that for any positive ε,

$$\Pr\left\{\left|\max_{t_0 \le t \le t_1} \frac{|U_m^{0,c}(t)|}{\{t(1-t)\}^{\frac{1}{2}}} - \max_{m_0 \le i \le m_1} \frac{|U_m^0(i/m)|}{\{(i/m)(1-i/m)\}^{\frac{1}{2}}}\right| > \varepsilon\right\} < \varepsilon.$$

By the definition of $U_m^{0,c}$, this is easily obtained. Then by (i),(ii),(iii), and (iv), the proof is completed. ∎

Corollary 2.2.3.

Under the same assumption as in Theorem 2.2.2,

$$M_2 = \max_{m_0 \le i \le m_1} \frac{|\tilde{U}_m^0(i/m)|}{\{(i/m)(1-i/m)\}^{\frac{1}{2}}} \to \max_{t_0 \le t \le t_1} \frac{|W^0(t)|}{\{t(1-t)\}^{\frac{1}{2}}} \qquad \text{in distribution.}$$

Proof : Since $\hat{\sigma}^2$ is a consistent estimator of σ^2 and $M_2 = M_1/\hat{\sigma}$, M_2 converges to the same limit as in (2.5) by the Slutsky's Lemma. ∎

Now we will consider the conditional test for H_0. This conditional test is based on the same test statistics, M_1 or M_2, but the rejection threshold depends on the x_j's, which are ancillary. In the following theorem, the asymptotic behavior of the MLRS will be considered conditionally on the x_j's when the x_j's are a random sample from some distribution.

Theorem 2.2.4.

Let $\mathbf{v}_j' = (x_j, y_j)$, $j = 1, \ldots, m$, be a sequence of i.i.d. random vectors such that $E[\mathbf{v}_j] = \boldsymbol{\mu}_j$ and $E[\mathbf{v}_j\mathbf{v}_j'] = \Sigma$.

Under H_0, as $m \to \infty$ and $m_i/m \to t_i$ for $i = 0, 1$, conditionally given $x_1, x_2, \cdots$,

$$M_1 = \max_{m_0 \le i \le m_1} \frac{|U_m^0(i/m)|}{\{(i/m)(1-i/m)\}^{\frac{1}{2}}} \to \max_{t_0 \le t \le t_1} \frac{|W^0(t)|}{\{t(1-t)\}^{\frac{1}{2}}} \qquad \text{in distribution,}$$

with probability 1.

Proof : This theorem is proved by basically the same argument as in the proof of Theorem 2.2.2. Note that

$$\frac{U_m^0(i/m)}{\{(i/m)(1-i/m)\}^{\frac{1}{2}}} = \frac{Z_m(i/m)}{\{(i/m)(1-i/m)\}^{\frac{1}{2}}} - \frac{R_m(i/m)}{\{(i/m)(1-i/m)\}^{\frac{1}{2}}},$$

where

$$S_i = \sum_{j=1}^{i}(y_j - \beta x_j - \alpha),$$

$$B_x = (\bar{x}_m - \bar{x}_i)i/\sqrt{m}\ , \quad \hat{\beta} = Q_{xy,m}/Q_{xx,m}\ ,$$

$$Z_m(i/m) = [S_i - S_m i/m]/[\sqrt{m}D_m(i/m, i/m)],$$

$$R_m(i/m) = (\beta - \hat{\beta})B_x(i/m)/D_m(i/m, i/m)$$

$$D_m(i/m, i/m) = 1 - [B_x(i/m)]^2/\{Q_{xx,m}(i/m)(1 - i/m)\}.$$

Then a.e. in x, as $m \to \infty$,

(i) $Z_m^c \to W^0$ in distribution,

(ii) $\max_{m_0 \le i \le m_1} R_m(i/m) \to 0$ in probability,

(iii) For any positive ε,

$$\Pr\left\{ \left| \max_{t_0 \le t \le t_1} \frac{|\ Z_m(t)\ |}{\{t(1-t)\}^{\frac{1}{2}}} - \max_{m_0 \le i \le m_1} \frac{|\ Z_m^c(i/m)\ |}{\{(i/m)(1-i/m)\}^{\frac{1}{2}}} \right| > \varepsilon \right\} < \varepsilon.$$

Combining these results, proof is completed. ∎

In proving Theorem 2.2.4, the necessary properties of the x_j's are

$$(\sum_{j=1}^{m} x_j)/m \to a \quad \text{and} \quad (\sum_{j=1}^{m} x_j^2)/m \to b \quad \text{a.e.}$$

In particular,

$$\bar{x}_m - \bar{x}_i \to 0 \quad \text{as} \quad m \to \infty \quad \text{and} \quad i \to \infty.$$

By the Theorems 2.2.2 and 2.2.4, it can be said that if the values of the independent variable are from some distribution, then the test statistic converges to the same limiting distribution whether we consider the test as the conditional or the unconditional one.

2.2.2. When the independent variable is fixed

In the previous section, we considered a case where (x_j, y_j) has a bivariate distribution such that $E[x_j^2] < \infty$ and $E[y_j^2] < \infty$. As a conditional test, we needed the convergence of the first and second moments of the independent variable to get the above limiting distribution. However, the independent variable is fixed in most applications and does not satisfy this condition in general. This section deals with asymptotic properties of the MLRS when the independent variable is fixed. We begin with the case in which the values of the independent variable are uniformly spaced. If $x_j = j/m$ for $j = 1, \ldots, m$, then $\bar{x}_m - \bar{x}_i = (1 - i/m)/2 \not\to 0$ any more and so the limiting distribution is not the same as in the previous section. First, we shall assume that σ^2 is known and hence without loss of generality equals one. Under the null hypothesis, we can write $U_m^0(i/m)$ as a weighted sum of ϵ_j's to prove Theorem 2.2.6 :

$$U_m^0(i/m) = \sum_{k=1}^{m} a_{i,k}\epsilon_k,$$

where

$$\begin{aligned} a_{i,k} &= \frac{i}{\{mD_m(i/m, i/m)\}^{\frac{1}{2}}}\left\{\frac{m-i}{mi} - \frac{(\bar{x}_i - \bar{x}_m)(x_k - \bar{x}_m)}{Q_{xx,m}}\right\}, & k \le i \\ &= \frac{i}{\{mD_m(i/m, i/m)\}^{\frac{1}{2}}}\left\{-\frac{1}{m} - \frac{(\bar{x}_i - \bar{x}_m)(x_k - \bar{x}_m)}{Q_{xx,m}}\right\}, & k > i, \end{aligned}$$

$$D_m(i/m, j/m) = 1 - (\bar{x}_m - \bar{x}_i)(\bar{x}_m - \bar{x}_j)mj/\{Q_{xx}(m-j)\} \quad \text{for } j > i. \tag{2.6}$$

Lemma 2.2.5.

Let $n \ge 2$ and $\{\mathbf{X_m} = (X_{1,m}, \ldots, X_{n,m})\}$ be a sequence of random elements of $\times_{k=1}^{n} \mathcal{C}_k$, (where $\mathcal{C}_k = \mathcal{C}[0,1]$) equipped with a product σ- field $\times_{k=1}^{n} \mathcal{B}_k$.

The sequence $\{\mathbf{X}_m\}$ is tight if and only if the n sets of marginal distributions,$\{X_{1,m}\}$, $\ldots$, $\{X_{n,m}\}$, are tight in $\mathcal{C}_1, \ldots, \mathcal{C}_n$.

Proof : Suppose that the sequence $\{\mathbf{X}_m = (X_{1,m}, \ldots, X_{n,m})\}$ is tight. Then there exists a compact set $\mathbf{K}$ in $\times_{k=1}^{n} \mathcal{B}_k$ such that

$$\Pr\{\mathbf{X}_j \in \mathbf{K}\} > 1 - \epsilon \quad \text{for all} \quad \mathbf{X}_j \in \{\mathbf{X}_m\}.$$

Let h_i be the mapping that carries the point $\mathbf{p} = (p_1, \ldots, p_n)$ in $\times_{i=1}^{n} \mathcal{B}_i$ to p_i in $\mathcal{B}_i$ for $i = 1, \ldots, n$. Since h_i is continuous for all i, $K' = h_i\mathbf{K}$ is compact and so $h_i^{-1}K' \supset \mathbf{K}$.

Then

$$\Pr\{X_{i,j} \in K'\} = \Pr\{\mathbf{X}_j \in h_i^{-1}K'\} > \Pr\{\mathbf{X}_j \in \mathbf{K}\} > 1 - \varepsilon,$$

which implies that the sequences, $\{X_{1,m}\}, \ldots, \{X_{n,m}\}$ are tight.

Conversely, suppose that $\{X_{1,m}\}, \ldots, \{X_{n,m}\}$ are tight sequences of random elements. Choose a positive ε. Then for each i, there exists a compact set K_i in $\mathcal{B}_i$ such that

$$\Pr\{X_{i,j} \in K_i\} > 1 - \varepsilon/n, \quad \text{for all } X_{i,j} \in \{X_{j,m}\}.$$

Let $\mathbf{K} = \bigcap_{i=1}^{n} h_i^{-1}K_i$. Then $\mathbf{K}$ is compact and

$$\Pr\{\mathbf{X}_j \in \mathbf{K}\} \geq 1 - \sum_{i=1}^{n} \Pr\{X_{i,j} \notin K_i\} > 1 - \varepsilon \ \text{ for all } \mathbf{X}_j \in \{\mathbf{X}_m\}.$$

Hence $\{\mathbf{X}_m\}$ is a tight sequence of random elements. ∎

Theorem 2.2.6.

Suppose that $x_j = j/m$ for $j = 1, \ldots, m$.

Under H_0, as $m \to \infty$ and $m_i/m \to t_i$ for $i = 0, 1$,

$$M_1 = \max_{m_0 \leq i \leq m_1} |\, U_m(i/m) \,| \to \max_{t_0 \leq t \leq t_1} |\, U(t) \,| \qquad \text{in distribution,} \tag{2.7}$$

where U is a Gaussian process with mean 0 and a covariance function,

$$\begin{aligned} \operatorname{Cov}[\, U(t), U(s) \,] &= \left\{\frac{t(1-s)}{s(1-t)}\right\}^{\frac{1}{2}} \frac{D(s,t)}{\{D(t,t)D(s,s)\}^{\frac{1}{2}}} \\ &= \sigma(t,s), \end{aligned}$$

where $D(s,t) = 1 - 3s(1-t)$ for $t < s$.

Proof : Recall that

$$U_m(i/m) = \frac{Z_m(i/m)}{\{(i/m)(1-i/m)\}^{\frac{1}{2}}} - \frac{R_m(i/m)}{\{(i/m)(1-i/m)\}^{\frac{1}{2}}},$$

where Z_m and R_m are defined in the proof of Theorem 2.2.4. By Theorem A 1.1, to show that $Z_m^c - R_m^c$ converges in distribution, we have only to check that the finite dimensional distributions converge in distribution and the sequence is tight. It can be easily shown that $\{Z_m^c\}$ and $\{R_m^c\}$ are tight. To prove the finite dimensional distributions of $Z_m^c - R_m^c$ converge to those of U^0 where U^0 is a Gaussian process with mean vector 0 and covariance matrix Λ to be defined later, we will show that for any sets of $(r_1, \ldots, r_n)$ and $(i_1, \ldots, i_n)$ such that $(i_1/m, \ldots, i_n/m) \to (t_1, \ldots, t_n)$ as $m \to \infty$,

$$E[\exp\{i\sum_{k=1}^{n} r_k(Z_m^c(i_k/m) - R_m^c(i_k/m))\}] \to E[\exp\{i\sum_{k=1}^{n} r_k U^0(t_k)\}] \text{ as } m \to \infty.$$

By using (2.6),

$$\begin{aligned} E[\exp\{i\sum_{k=1}^{n} r_k(Z_m^c(i_k/m) - R_m^c(i_k/m))\}] &= E[\exp\{i\sum_{k=1}^{n}\sum_{j=1}^{m} r_k a_{i_k,j}\varepsilon_j\}] \\ &= E[\exp i(\mathbf{b}'\boldsymbol{\varepsilon})], \end{aligned}$$

where $\mathbf{b} = (b_1, \ldots, b_m)$ with $b_j = \sum_{k=1}^{n} r_k a_{i_k,j}$, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_m)$. Now elementary algebra shows that

$$\sum_{j=1}^{m} b_j^2 \to \mathbf{r}'\Lambda\mathbf{r},$$

where $\mathbf{r}' = (r_1, \ldots, r_n)$, and Λ is a matrix whose (k,l)-th entry is $\sigma(t_k, t_l)\{t_k(1-t_k)t_l(1-t_l)\}^{\frac{1}{2}}$. Hence this implies that

$$\begin{aligned} &(Z_m^c(i_1/m) - R_m^c(i_1/m), \ldots, Z_m^c(i_n/m) - R_m^c(i_n/m)) \\ &\qquad \to (U^0(t_1), \ldots, U^0(t_n)) \qquad \text{in distribution.} \end{aligned}$$

Now tightness of the sequence $\{Z_m^c - R_m^c\}$ follows from Lemma 2.2.5 and Lemma 7 (Billingsley, 1968). Thus $Z_m^c - R_m^c$ converges to U^0 in distribution.

By the continuous mapping theorem,

$$M_1 \to \max_{t_0 \le t \le t_1} \frac{|\, U^0(t) \,|}{\{t(1-t)\}^{\frac{1}{2}}} \qquad \text{in distribution.}$$

It is evident that $U^0(t)/\{t(1-t)\}^{\frac{1}{2}}$ is a Gaussian process. Since a Gaussian process is completely determined by mean vector and covariance matrix, (2.7) holds. ∎

Now we will generalize this result to the case where $x_j = f(j/m)$ for some integrable function f.

Lemma 2.2.7.

Suppose that $x_j = f(j/m)$ $j = 1, \ldots, m$, for some integrable function f such that $f(0) = 0$ and $f(1) = 1$.

Then for $i < j$, as $i/m \to t$, $j/m \to s$, $m \to \infty$,

$$\operatorname{Cov}[U_m(i/m), U_m(j/m)] = \left\{\frac{(i/m)(1-j/m)}{(j/m)(1-i/m)}\right\}^{\frac{1}{2}} \frac{D_m(i/m, j/m)}{\{D_m(i/m, i/m) D_m(j/m, j/m)\}^{\frac{1}{2}}}$$

where

$$g_m(i/m) = (\bar{x}_m - \bar{x}_i)/\{(1 - i/m)\sqrt{Q_{xx}/m}\},$$

$$D_m(i/m, j/m) = 1 - (\bar{x}_m - \bar{x}_i)(\bar{x}_m - \bar{x}_j)mj/\{Q_{xx}(m-j)\}$$

$$= 1 - (j/m)(1 - i/m) g_m(i/m) g_m(j/m) \qquad \text{for } j > i.$$

$$\to \left\{\frac{t(1-s)}{s(1-t)}\right\}^{\frac{1}{2}} \frac{D(t,s)}{\{D(t,t)D(s,s)\}^{\frac{1}{2}}}. \tag{2.8}$$

where

$$g(t) = \frac{\int_0^1 f(u)du - [\int_0^t f(u)du]/t}{(1-t)\{\int_0^1 f^2(u)du - [\int_0^1 f(u)du]^2\}^{\frac{1}{2}}}$$

$$D(t,s) = 1 - s(1-t)g(t)g(s) \qquad \text{for } s > t.$$

Proof : First we will derive $\operatorname{Cov}[U_m(i/m), U_m(j/m)]$ for $i < j$. Note that $U_m(i/m)$ can be written as

$$[(\bar{y}_m - \bar{y}_i) - (\bar{x}_m - \bar{x}_i)Q_{xy,m}/Q_{xx,m}]/\{D_m(i/m, i/m)(m-i)/(mi)\}^{\frac{1}{2}}.$$

Assuming the ϵ_j's are normally distributed with mean 0 and variance 1, it can be easily shown that

$$\begin{pmatrix} \bar{y}_m - \bar{y}_i \\ (\bar{x}_m - \bar{x}_i)\hat{\beta} \end{pmatrix} \sim N\left(\begin{pmatrix} \beta(\bar{x}_m - \bar{x}_i) \\ \beta(\bar{x}_m - \bar{x}_i) \end{pmatrix}, \begin{pmatrix} (m-i)/(mi) & (\bar{x}_m - \bar{x}_i)/Q_{xx,m} \\ (\bar{x}_m - \bar{x}_i)/Q_{xx,m} & (\bar{x}_m - \bar{x}_i)^2/Q_{xx,m} \end{pmatrix} \right).$$

where $\hat{\beta} = Q_{xy,m}/Q_{xx,m}$.

Also it is straightforward to show

$$\operatorname{Cov}[\bar{y}_m - \bar{y}_i, \bar{y}_m - \bar{y}_j] = (m-j)/(mj).$$

Thus

$$\operatorname{Cov}[U_m(i/m), U_m(j/m)] = \left\{\frac{(i/m)(1-j/m)}{(j/m)(1-i/m)}\right\}^{\frac{1}{2}} \frac{D_m(i/m, j/m)}{\{D_m(i/m, i/m)D_m(j/m, j/m)\}^{\frac{1}{2}}}$$

Since

$$\bar{x}_m - \bar{x}_i = \sum_{k=1}^{m} f(k/m)/m - \sum_{k=1}^{i} f(k/m)/i$$

and

$$Q_{xx,m} = \sum_{k=1}^{m} f^2(k/m) - \Big[\sum_{k=1}^{m} f(k/m)\Big]^2 /m,$$

it can be easily shown that $g_m(i/m) \to g(t)$ as $m \to \infty$, $i \to \infty$. Then (2.8) follows immediately. ∎

Lemma 2.2.7 says that the test statistic depends on the x_j's only through the function, g_m. When $x_j = j/m$, $g_m(j/m) = \sqrt{3}$ for all j. The same argument as in the proof of Theorem 2.2.6 leads to the following theorem.

Theorem 2.2.8.

Suppose that $x_j = f(j/m)$ $j = 1, \ldots, m$, for some integrable function f such that $f(0) = 0$ and $f(1) = 1$.

Under H_0, as $m \to \infty$ and $m_i/m \to t_i$ for $i = 0, 1$,

$$M_1 = \max_{m_0 \le i \le m_1} |\, U_m(i/m)\,| \to \max_{t_0 \le t \le t_1} |\, U(t)\,| \qquad \text{in distribution,} \tag{2.9}$$

where U is a Gaussian process with mean 0 and a covariance function

$$\begin{aligned}\operatorname{Cov}[\,U(t), U(s)\,] &= \left\{\frac{t(1-s)}{s(1-t)}\right\}^{\frac{1}{2}} \frac{D(s,t)}{\{D(t,t)D(s,s)\}^{\frac{1}{2}}} \\ &= \sigma(t,s),\end{aligned}$$

where $D(s,t) = 1 - s(1-t)g(t)g(s)$ for $t < s$.

Corollary 2.2.9.

Under the same assumption as in Theorem 2.2.8,

$$M_2 = \max_{m_0 \le i \le m_1} |\, \tilde{U}_m(i/m)\,| \to \max_{t_0 \le t \le t_1} |\, U(t)\,| \qquad \text{in distribution.}$$

Remark 2.1. Comparing with the case in which the independent variable is random, we see that $D(t,s)/\{D(t,t)D(s,s)\}^{\frac{1}{2}}$ is an additional factor in the covariance of the limiting process. Since the mean and variance remain the same, the configuration of the values of the independent variable affect the distribution of the MLRS only through this additional term in the covariance function.

Remark 2.2. Since the limiting distribution of the MLRS in (2.5) involves Brownian motion, the stochastic process in that limit has non-differentiable sample paths. In the case of the fixed independent variable, U in (2.7) also has non-differentiable sample paths.

2.3. Approximations to Significance Levels

As described in Section 2.1, the exact distributions of the test statistics are quite difficult to analyse. In this section, we give approximations to the right-hand tail of the null distributions of M_1 and M_2 and perform Monte Carlo experiments to see how accurate these approximations are. As in Section 2.2, we first consider the case in which the x_j's are random and then the case where the x_j's are fixed. In both cases, we study the significance levels as boundary crossing probabilities by discrete stochastic processes with nondifferentiable sample paths. Approximations to significance levels are derived for the MLRS with known variance, M_1, and will be discussed how well this can be applied to the unknown variance case.

2.3.1. When the independent variable is random

When the independent variable is random, Theorem 2.2.2 shows that the MLRS, M_1, converges to

$$\max_{t_0 \le t \le t_1} \frac{|W^0(t)|}{\{t(1-t)\}^{\frac{1}{2}}}, \qquad \textit{in distribution},$$

where W^0 is a Brownian bridge process on $[0,1]$ and $m_i/m \to t_i$, for $i = 0, 1$, as $m \to \infty$. Thus the significance level of the test, $\Pr\{M_1 > b\}$, can be approximated by that of this limiting distribution. Siegmund(1986) provides the approximation to

$$\Pr\{\max_{t_0 \le t \le t_1} |W^0(t)|/\{t(1-t)\}^{\frac{1}{2}} \ge b\}.$$

which is quite general since it can be applied also if the underlying distribution of the observations is not normal. However Table 1 shows that this approximation overestimates the actual values about 200 %. In Table 1 and in other tables, we obtained the percentiles by a Monte Carlo experiment using simple random sampling with 10,000 samples for each situation. As a correction for discrete time, (19) in James, James, and Siegmund (1987) was used and that result is also summarized in Table 1. This discrete approximation does not perform perfectly but it gives a rough idea about the significance level.

Table 2 concerns the MLRS with unknown variance, M_2. Using (21) in James, James, and Siegmund (1987), a similar kind of result is obtained. The numbers in parenthesis are the approximations to the significance levels of M_2 using the approximation derived for the known variance case. This gives some insight about whether the approximation derived for the known variance case can be applied to the unknown variance case. Since in the next section we will derive an approximation to the significance level of M_1 when the independent variable is fixed and see how that works for the unknown variance case, we will discuss this more later.

2.3.2. When the independent variable is fixed

As we can see in Theorem 2.2.6 and Theorem 2.2.8, the limiting distribution is not a function of a Brownian motion but involves a different Gaussian process when the independent variable is fixed. In this section, in order to get an approximation to the significance level of M_1, we begin with the case where $x_j = j/m$ and later consider more general configurations of the independent variable. In principle, Durbin (1985) gave approximation formula to the probability of boundary crossing by a continuous Gaussian process satisfying some conditions. However as before these are not accurate since these did not take discreteness into consideration.

The main result of this section is a new approximation taking discreteness into account. Assuming the normality of the error variable, we can consider the significance level of M_1 as a boundary crossing probability by the Gaussian process, U_m, defined on $\{i : m_0 \leq i \leq m_1\}$. As discussed before, our Gaussian process is nonstationary and nondifferentiable. To approximate the boundary crossing probability by the discrete stochastic process whose limiting process has a non-differentiable sample path, the argument in Leadbetter, Lindgren and Rootzen(1983, Chapter 12), as modified for discrete time by Hogan and Siegmund (1986), will be used. We start with the given discrete time Gaussian process and derive an approximation to the boundary crossing probability by this discrete process as the sample size gets large. In Leadbetter, Lindgren and Rootzen (1983, Chapter12), their goal is to approximate the boundary crossing probability by a non-differentiable continuous Gaussian process. They considered the probability of crossing the boundary by the given process at discrete instants of time first and let the interval of each time points get smaller and smaller . Actually we get the same result if we consider the continuous limiting process and find an asymptotic expression for the boundary crossing probability by this limiting process observed only at the discrete instants of time. From Lemma 2.3.1 through Theorem 2.3.5, it is assumed $x_j = j/m$ for $j = 1, \ldots, m$, and to obtain nontrivial limits as $b \to \infty$, we use the normalized process, $U^t_{b,m}(i) = b(U_m(t + i/m) - b)$, where $b^2/m \to a$. In order to state approximations to the significance levels in Theorems 2.3.5 and 2.3.7, it is helpful to introduce the function

$$\nu(x) = 2x^{-2}\exp\{-2\sum_{n=1}^{\infty} n^{-1}\Phi(-\frac{1}{2}xn^{\frac{1}{2}})\}, \quad (x > 0) \tag{2.10}$$

where Φ denotes the standard normal distribution function. The function ν was used by Siegmund (1985) and is easily evaluated numerically by (2.10) or approximately as suggested in Siegmund (1985, ChX).

Lemma 2.3.1.

Suppose that $x_j = j/m$ for $j = 1, \ldots, m$.

Let $U^t_{b,m}(i) = b(U_m(t + i/m) - b)$, and suppose $m \to \infty$, $b \to \infty$ so that $b^2/m \to a$.

Then, the conditional distributions of $U^t_{b,m}(i)$ given that $U^t_{b,m}(0) = x$ are normal with

$$E\,[U^t_{b,m}(i) - x | U^t_{b,m}(0) = x] = -\mu_a(t)i + o(1), \tag{2.11}$$

$$\mathrm{Cov}\,[U^t_{b,m}(i) - x,\ U^t_{b,m}(j) - x | U^b_{t,m}(0) = x] = 2\mu_a(t)\min(i,j) + o(1). \tag{2.12}$$

where $\mu_a(t) = a/[2t(1-t)\{1-3t(1-t)\}]$.

Proof : Using Taylor series expansion of covariance functions and doing a tedious calculation, (2.11) and (2.12) are obtained. ▮

The first step in our derivation to the distribution of M_1 is to consider the maximum taken over a fixed number of points, $t, t+1/m, \ldots, t+n/m$.

Lemma 2.3.2.

For fixed n and a, as $b \to \infty$ and $m \to \infty$,

$$\Pr\{\max_{0\le i\le n} U_m(t+i/m) \ge b\}/[\frac{\phi(b)}{b}] \to 1 + H_a(t,n), \qquad (2.13)$$

where

$$H_a(t,n) = \int_{-\infty}^{0} \exp(-x)\Pr\{\max_{0\le i\le n} Y_a^t(i) \ge x\}dx,$$

and $Y_a^t(i)$ is a partial sum of i.i.d. random variables with

$$Y_a^t(1) \sim N(-\mu_a(t), 2\mu_a(t)).$$

Proof : Since the conditional distribution is normal, it is determined by its mean and covariance. Then the previous lemma implies the limiting process can be represented by

$$Y_a^t(i) = \sigma_a(t)W(i) - \mu_a(t)i,$$

where W is a standard Brownian motion and $\sigma_a^2(t) = 2\mu_a(t)$.

Then, following the same argument in Lemma 12.2.3 of Leadbetter, Lindgren, and Rootzen (1983), (2.13) holds. ▮

Lemma 2.3.3.

There exists a function $H_a^*(t)$ such that

$$\lim_{n\to\infty} H_a(t,n)/n = H_a^*(t) \qquad \text{uniformly in t.}$$

As $b \to \infty$ and $m \to \infty$,

$$\Pr\{\max_{t_0 \le i/m \le t_1} U_m(i/m) \ge b\} / [b\phi(b)] \to \int_{t_0}^{t_1} H_a^*(t)dt/a. \tag{2.14}$$

Proof : Let

$$\begin{aligned} B_k &= \{\max_{kn \le i \le (k+1)n} U_m(i/m) \ge b\} \\ &= \{\max_{0 \le i \le n} U_m(kn/m + i/m) \ge b\} \end{aligned}$$

Then it can be shown that

$$\Pr\{\max_{t_0 \le i/m \le t_1} U_m(i/m) \ge b\} \sim \sum_{k=K_0}^{K_1} P\{B_k\},$$

where $\lfloor K_0 n \rfloor = m_0$, $\lfloor K_1 n \rfloor = m_1$, and $\lfloor x \rfloor$ denotes the greatest integer which is less than x.

By Lemma 2.3.2,

$$\begin{aligned} \sum_{k=K_0}^{K_1} P\{B_k\} &\sim [\phi(b)/b] \sum_{k=K_0}^{K_1} \left[1 + H_a(kn/m, n)\right] \\ &\sim b\phi(b)\left[1/na + \sum_{k=K_0}^{K_1} H_a(kn/m, n)/b^2\right]. \end{aligned}$$

And thus

$$\begin{aligned} &\Pr\{\max_{t_0 \le i/m \le t_1} U_m(i/m) \ge b\}/[b\phi(b)] \\ &\sim 1/na + \sum_{k=K_0}^{K_1} H_a(kn/m, n)/b^2 \\ &\sim 1/na + \int_{t_0}^{t_1} H_a(t, n)dt/(na) \end{aligned}$$

The proof is completed by letting $n \to \infty$ and proceeding as in Lemma 12.2.4 of Leadbetter, Lindgren, and Rootzen (1983). ∎

The last step is to evaluate H_a^* in (2.14). In evaluating H_a^*, we use the argument in Siegmund (1985, Ch VIII), which leads to the derivation of the boundary crossing probability by a random walk with unit variance.

Lemma 2.3.4.

$$\int_{t_0}^{t_1} H_a^*(t)dt/a = \int_{t_0}^{t_1} \mu_a(t)\nu[2\mu_a^*(t)]\, dt\ /a, \tag{2.15}$$

where

$$\mu_a(t) = a/[2t(1-t)\{1-3t(1-t)\}],$$

$$\mu_a^*(t) = \{\mu_a(t)/2\}^{\frac{1}{2}}.$$

Proof : Note that

$$H_a(t,n) = \int_{-\infty}^{0} \exp(-x)\Pr\{\max_{0\le i\le n} Y_a^t(i) \ge x\}dx,$$

where $Y_a^t(i)$ is a partial sum of i-i.i.d.random variables defined in Lemma 2.3.2. Let $Y_a^{t,*}(1) = Y_a^t(1)/\sigma_a(t)$ to make the variance equal to 1.

Then the Wald's likelihood ratio identity implies that

$$H_a(t,n) = \int_0^{\infty} \exp[y\sigma_a(t)]\Pr\{\max_{0\le i\le n} Y_a^{t,*}(i) \ge y\}dy,$$

$$= \sigma_a(t)\int_0^{\infty} E_{\mu^*}[\exp\{-2\mu_a^*(t)R_y\} : T_y \le n]dy,$$

where

$$T_y = inf\{n \ge 1 : Y_a^{t,*}(n) \ge y\},$$

$$R_y = Y_a^{t,*}(T_y) - y.$$

Hence it suffices to evaluate the limit as $n \to \infty$ of

$$n^{-1}\int_0^{\infty} E_{\mu^*}[\exp\{-2\mu_a^*(t)R_y\} : T_y \le n]dy.$$

By the same argument in Lemma 3.4 of Hogan and Siegmund (1986), this is approximated by $\mu_a^*(t)\nu[2\mu_a^*(t)]$, as $n \to \infty$. Therefore (2.15) holds. ∎

By combining Lemmas 2.3.1, 2.3.2, 2.3.3, and 2.3.4, we obtain the following approximation to the tail of the distribution of the maximum, $M_1 = \max_{m_0\le i\le m_1} U_m(i/m)$, over an interval $[m_0, m_1]$.

Theorem 2.3.5.

Assume that $b \to \infty$, $m_0 \to \infty$, $m_1 \to \infty$, and $m \to \infty$ in such a way that for some $0 < t_0 < t_1 < 1$ and $a > 0$

$$m_i/m \to t_i, \quad i = 0, 1 \quad \text{and } b^2/m \to a.$$

Then as $m \to \infty$,

$$Pr\{\max_{m_0 \le i \le m_1} U_m(i/m) \ge b\} \sim b\phi(b) \int_{t_0}^{t_1} \nu[2\mu_a^*(t)]\mu_a(t)dt/a,$$

where

$$\mu_a(t) = a/[t(1-t)\{1 - 3t(1-t)\}],$$
$$\mu_a^*(t) = \{\mu_a(t)/2\}^{\frac{1}{2}}/2.$$

Remark 2.3. When $x_j = j/m$, $j = 1, \ldots, m$, the significance level of the test can be approximated by

$$\Pr\{\max_{m_0 \le i \le m_1} |U_m(i/m)| \ge b\} \sim 2\Pr\{\max_{m_0 \le i \le m_1} U_m(i/m) \ge b\}$$
$$\sim b\phi(b) \int_{t_0}^{t_1} 2\mu_a(t)\nu[2\mu_a^*(t)]dt/a. \qquad (2.16)$$

Table 3 gives an indication of the accuracy of (2.16). As before, percentiles of M_1, b_1, were obtained by the same kind of Monte Carlo experiment. Table 3 also indicates that the approximation (2.16) can be applied to the unknown variance case. In Table 3, b_2 are the percentiles of M_2 for various sample sizes and it can be said that approximations are reasonably accurate if sample sizes are big enough and $\alpha \le 0.1$. Since the case of $x_j = j/m$ can be applied to the regression model in which x are equally spaced time points, which arises often in statistical analysis, we provide in Table 4 the tail probabilities of M_1 when $x_j = j/m$ under H_0.

In the remaining part of this section, an approximation to the significance level for a general configuration of the values of the x_j's will be derived and numerical results will be presented. Proofs will be omitted since they follow closely those of the previous theorem.

Lemma 2.3.6.

Suppose that $x_j = f(j/m)$, $j = 1, \ldots, m$, for some integrable function f such that $f(0) = 0$ and $f(1) = 1$.

Then as $m \to \infty$ and $b \to \infty$ in such a way that $b^2/m \to a$,

$$E[U^t_{b,m}(i) - x | U^t_{b,m}(0) = x] \to -\mu_a(t)i, \tag{2.17}$$

$$\text{Cov}\ [U^t_{b,m}(i) - x,\ U^t_{b,m}(j) - x | U^b_{t,m}(0) = x] \to 2\mu_a(t)\min(i,j), \tag{2.18}$$

where $\mu_a(t) = a/[2t(1-t)\{1 - g^2(t)t(1-t)\}]$, and g was defined in Lemma 2.2.6.

Proof : (2.17) and (2.18) directly follow from a long calculation. ∎

Theorem 2.3.7.

Suppose that $x_j = f(j/m)$, $j = 1, \ldots, m$, for some integrable function f such that $f(0) = 0$ and $f(1) = 1$.

Assume that $b \to \infty$, $m_0 \to \infty$, $m_1 \to \infty$, and $m \to \infty$ in such a way that for some $0 < t_0 < t_1 < 1$ and $a > 0$

$$m_i/m \to t_i, \quad i = 0.1 \quad \text{and } b^2/m \to a.$$

Then as $m \to \infty$,

$$\begin{aligned} Pr\{ \max_{m_0 \le i \le m_1} |U_m(i/m)| \geq b\} \\ \sim b\phi(b) \int_{t_0}^{t_1} \nu[2\mu^*_a(t)]/[t(1-t)\{1 - g^2(t)t(1-t)\}]dt, \end{aligned} \tag{2.19}$$

where $\mu^*_a(t) = \{a/[t(1-t)\{1 - g^2(t)t(1-t)\}]\}^{\frac{1}{2}}/2$.

Table 5 supports that the theoretical approximation (2.19) is quite accurate when $x_j = (j/m)^2$. In Table 5, p_1 is obtained for $f(x) = x^2$. and p_2 is the approximation using the linear interpolation of the x_j's as f. We get the percentiles for unknown variance case by Monte Carlo method and approximates significance levels using the approximation formula derived for known variance case. Even though they are not perfect, a rough idea about the tail probability can be obtained from them.

2.4. Powers and Confidence Regions

In this section we follow the arguments of James, James, and Siegmund(1987) to obtain an approximation to the power of (2.1). We derive an approximation of the power of (2.1) for the fixed x_j's, starting from the uniformly spaced x_j's. Suppose that we observe $y_1, \ldots, y_m$ as in section 2.2 and $x_j = j/m$ for $j = 1, \ldots, m$, that there is exactly one change point, ρ, only in the intercept term of the regression line and that α_1, α_2, and β are unknown parameters. In order to get an intuitive idea of the boundary crossing by the given stochastic process, we consider a modified stochastic process and a curved boundary as follows. Let $U_m^*(i) = U_m(i/m)\{i(1-i/m)\}^{\frac{1}{2}}$. Then from (2.2) and (2.3) it can be easily seen that the process $U_m^*(i)$ $(i = m_0, \ldots, m_1)$ has the mean value,

$$\begin{aligned} E[U_m^*(i)] &= i(1-\rho/m)\frac{D_m(i/m, \rho/m)}{\{D_m(i/m, i/m)\}^{\frac{1}{2}}}(\alpha_2 - \alpha_1), \quad i \le \rho \\ &= \rho(1-i/m)\frac{D_m(i/m, \rho/m)}{\{D_m(i/m, i/m)\}^{\frac{1}{2}}}(\alpha_2 - \alpha_1), \quad i > \rho, \end{aligned} \tag{2.20}$$

and the covariance function for $i < j$,

$$\mathrm{Cov}\,[U_m^*(i), U_m^*(j)] = i(1-j/m)r_m(i,j) \tag{2.21}$$

where

$$D_m(i/m, j/m) = 1 - (\bar{x}_m - \bar{x}_i)(\bar{x}_m - \bar{x}_j)mj/\{Q_{xx,m}(m-j)\} \quad \text{for } i < j,$$
$$r_m(i,j) = D_m(i/m, j/m)/\{D_m(i/m, i/m)D_m(j/m, j/m)\}^{\frac{1}{2}}.$$

For $1 \le m_0 < m_1 < m$, let

$$\begin{aligned} T_0 &= \inf\{i : i \ge m_0,\ |U_m^*(i)| \ge b\{i(1-i/m)\}^{\frac{1}{2}}\}, \\ T_1 &= \sup\{i : i \le m_1,\ |U_m^*(i)| \ge b\{i(1-i/m)\}^{\frac{1}{2}}\}, \end{aligned} \tag{2.22}$$

and let $\Pr_\xi^{(\rho)}\{T_0 \le m_1\} = \Pr\{T_0 \le m_1 \mid U_m^*(\rho) = \xi\}$. The power of the test defined by (2.1) is of the form, $\Pr_\rho\{M_1 \ge b\} = \Pr_\rho\{T_0 \le m_1\}$, where $m_0 < \rho < m_1$. It is obvious that

$$\begin{aligned} \Pr_\rho\{T_0 \le m_1\} = \Pr\{\,|U_m^*(\rho)| \ge b\{\rho(1-\rho/m)\}^{\frac{1}{2}}\} \\ + \Pr\{\,|U_m^*(\rho)| < b\{\rho(1-\rho/m)\}^{\frac{1}{2}},\ T_0 \le m_1\} \end{aligned}$$

$$= \Pr\{\, |U_m^*(\rho)| \geq b\{\rho(1-\rho/m)\}^{\frac{1}{2}}\}$$
$$+ \int_{-b\{\rho(1-\rho/m)\}^{\frac{1}{2}}}^{b\{\rho(1-\rho/m)\}^{\frac{1}{2}}} \Pr_\xi^{(\rho)}\{T_0 < m_1\} \Pr\{U_m^*(\rho) \in d\xi\} \qquad (2.23)$$

Since the marginal distribution of $U_m^*(i)$ is known, to approximate (2.23) it suffices to approximate the conditional probability in (2.23). To approximate $\Pr_\xi^{(\rho)}\{T_0 < m_1\}$ we may assume that

$$|\xi| = b\{\rho(1-\rho/m)\}^{\frac{1}{2}} - x \qquad (2.24)$$

with $x = O(1)$ as $m \to \infty$, since the principal contribution to the integral on the right-hand side of (2.23) comes from values of ξ close to the boundary value. Given $U_m^*(\rho) = \zeta$ of the form (2.24), if $|U_m^*(i)| \geq b\{i(1-i/m)\}^{\frac{1}{2}}$ for some $m_0 \leq i < \rho$ and $|U_m^*(j)| \geq b\{j(1-j/m)\}^{\frac{1}{2}}$ for some $\rho < j \leq m_1$, this event with overwhelming probability occurs for some i and j which are closed to ρ. Moreover, given $U_m^*(\rho) = \xi$, asymptotically as $m \to \infty$ the processes $U_m^*(i)$ $(i = m_0, \ldots, \rho)$ and $U_m^*(j)$ $(j = \rho+1, \ldots, m_1)$ are conditionally independent for i and j close to ρ. Thus we can write

$$\Pr_\xi^{(\rho)}\{T_0 < m_1\} \cong \Pr_\xi^{(\rho)}\{T_0 < \rho\} + \Pr_\xi^{(\rho)}\{T_1 > \rho\} - \Pr_\xi^{(\rho)}\{T_0 < \rho\} \Pr_\xi^{(\rho)}\{T_1 < \rho\}. \quad (2.25)$$

Since both probabilities on the right hand side of (2.25) are of the same form, it is enough to consider the first one. To approximate the first probability, we assume that m is large and that ρ and $\rho - m_0$ are proportional to m.

Lemma 2.4.1.

Let $\xi = b\{\rho(1-\rho/m)\}^{\frac{1}{2}} - x = m\xi_0$. For $i \ll \rho$, given $U_m^*(\rho) = \xi$,

as $m \to \infty$, $\rho/m \to t^*$, for each fixed i,

$$[\, U_m^*(\rho - i) - \mu(\rho - i\,|\,\rho)\,]\{D(t^*, t^*)\}^{\frac{1}{2}}$$

is distributed approximately as $S_i = \sum_{k=1}^{i} z_k$, where z_k's are i.i.d. standard normal random variables and $\mu(\rho - i\,|\,\rho) = E[U_m^*(\rho - i)\,|\,U_m^*(\rho) = \xi]$,

where $D(t^*, t^*) = 1 - 3t^*(1 - t^*)$.

Proof : Since $U_m^*(i)$ is distributed as $N(\mu(i|\rho), \sigma^2(i|\rho))$ given $U_m^*(\rho) = \xi$, where

$$\mu(i|\rho) = \xi r_m(i,\rho)i/\rho,$$
$$\sigma^2(i|\rho) = i(1 - i/m) - [r_m(i,\rho)]^2 i^2(1 - \rho/m)/\rho,$$

it can be easily obtained that as $m \to \infty$,

$$\mathrm{Cov}\,[U_m^*(\rho - i),\ U_m^*(\rho - j) \,|\, U_m^*(\rho) = \xi] \to min(i,j)/D(t^*,t^*).$$

Thus $[U_m^*(\rho - i) - \mu(\rho - i \,|\, \rho)]\{D(t^*,t^*)\}^{\frac{1}{2}}$ behaves like a sum of independent normally distributed random variables, each having mean 0 and variance 1. ∎

Now we define stopping times τ_0^+ and τ_0^- as follows :

$$\tau_0^+ = \inf\{i : i \geq m_0,\ U_m^*(i) \geq b\{i(1 - i/m)\}^{\frac{1}{2}}\},$$
$$\tau_0^- = \inf\{i : i \geq m_0,\ U_m^*(i) \leq -b\{i(1 - i/m)\}^{\frac{1}{2}}\}.$$

Lemma 2.4.2.

Suppose that $b \to \infty$, $\rho \to \infty$, $m \to \infty$ in such a way that $b/\sqrt{m} \to b_0$, and $\rho/m \to t^*$. Let $x = b\{\rho(1 - \rho/m)\}^{\frac{1}{2}} - \xi = O(1)$.

Then as $m \to \infty$,

$$\mathrm{Pr}_\xi^{(\rho)}\{\tau_0^+ < \rho\} \cong \nu[2\eta]\exp[-2\eta x\{D(t^*,t^*)\}^{\frac{1}{2}}],$$

and similarly

$$\mathrm{Pr}_{-\xi}^{(\rho)}\{\tau_0^- < \rho\} \cong \nu[2\eta]\exp[-2\eta x\{D(t^*,t^*)\}^{\frac{1}{2}}], \tag{2.26}$$

where $\eta = b_0/[2\{D(t^*,t^*)t^*(1 - t^*)\}^{\frac{1}{2}}]$.

Proof : By (2.20) and elementary calculus, it can be seen that for fixed x, i,

$$b\{(\rho - i)[1 - (\rho - i)/m]\}^{\frac{1}{2}} - \mu(\rho - i \,|\, \rho) \to x + ib_0/[2D(t^*,t^*)\{t^*(1 - t^*)\}^{\frac{1}{2}}].$$

From Lemmas 2.4.1,

$$\Pr_{\xi}^{(\rho)}\{\tau_0^+ < \rho\} = \Pr_{\xi}^{(\rho)}\{U_m^*(\rho - i) - \mu(\rho - i|\rho) > b\{(\rho - i)(1 - (\rho - i)/m)\}^{\frac{1}{2}} - \mu(\rho - i|\rho) \quad \text{for some } 1 \le i < \rho - m_0\}$$
$$\to \Pr_0\{S_i > i\eta + x\{D(t^*, t^*)\}^{\frac{1}{2}} \quad \text{for some } i \ge 1\},$$

where $\eta = b_0/[2\{D(t^*, t^*)t^*(1 - t^*)\}^{\frac{1}{2}}]$ and S_i was given in Lemma 2.4.1. Therefore this conditional probability is approximately the same as

$$\Pr_{-\eta}\{S_i^* > y \quad \text{for some } i \ge 1\},$$

where S_i^* is a partial sum of the i.i.d.random variables, each having mean $-\eta$ and variance 1 and $y = x\{D(t^*, t^*)\}^{\frac{1}{2}}$. Following the argument in Siegmund (1986, Ch VIII), this probability can be approximated by $\nu(2\eta)\exp[-2\eta x\{D(t^*, t^*)\}^{\frac{1}{2}}]$, which can be used as an approximation to the conditional probability in (2.23). ∎

Theorem 2.4.3.

Suppose that $b \to \infty$, $\rho \to \infty$, $m \to \infty$ in such a way that $b/\sqrt{m} \to b_0$, and $\rho/m \to t^*$. Then as $m \to \infty$,

$$\Pr_{\rho}\{M_1 > b\} \sim [1 - \Phi(\gamma)] + m^{-\frac{1}{2}}\phi(\gamma)\left[\frac{2\nu(2\eta)}{\delta\{t^*(1 - t^*)D(t^*, t^*)\}^{\frac{1}{2}}} - \frac{\nu^2(2\eta)}{m^{\frac{1}{2}}(b_0 + \delta\{t^*(1 - t^*)D(t^*, t^*)\}^{\frac{1}{2}})}\right], \tag{2.27}$$

where

$$\gamma = m^{\frac{1}{2}}[b_0 - \delta\{t^*(1 - t^*)D(t^*, t^*)\}^{\frac{1}{2}}]$$

and D and η are given in Lemma 2.4.2.

Proof: Note that

$$\Pr_{\xi}^{(\rho)}\{T_0 < \rho\} \cong \Pr_{\xi}^{(\rho)}\{\tau_0^+ < \rho\} + \Pr_{\xi}^{(\rho)}\{\tau_0^- < \rho\} - \Pr_{\xi}^{(\rho)}\{\tau_0^+ < \rho\}\Pr_{\xi}^{(\rho)}\{\tau_0^- < \rho\}.$$

Using Lemma 2.4.2 and the fact that the major contribution to $\int_{-b\{\rho(1-\rho/m)\}^{\frac{1}{2}}}^{b\{\rho(1-\rho/m)\}^{\frac{1}{2}}} \Pr_{\xi}^{(\rho)}\{T_0 < \rho\}\Pr\{U_m^*(\rho) \in d\xi\}$ comes from the probability of crossing the upper boundary by the process $U_m^*(i)$ conditioned on $U_m^*(\rho) = \xi$ which is close to the upper boundary, we get

$$\int_{-b\{\rho(1-\rho/m)\}^{\frac{1}{2}}}^{b\{\rho(1-\rho/m)\}^{\frac{1}{2}}} \mathrm{Pr}_\xi^{(\rho)}\{T_0 < \rho\} \, \mathrm{Pr}\{U_m^*(\rho) \in d\xi\}$$
$$\sim \int_{-\infty}^{b\{\rho(1-\rho/m)\}^{\frac{1}{2}}} \exp[-2\eta\{b\{\rho(1-\rho/m)\}^{\frac{1}{2}} - \xi\}\{D(t^*,t^*)\}^{\frac{1}{2}}] \, \mathrm{Pr}\{U_m^*(\rho) \in d\xi\}$$

Using a similar approximation for the other conditional probability on the right-hand side of (2.27), and evaluating the integral in (2.22) asymptotically as $m \to \infty$, we obtain (2.25).

∎

Remark 2.4. When $x_j = f(j/m)$ for some integrable function f such that $f(0) = 0$ and $f(1) = 1$, we get the same result but with $D(t^*,t^*)$ defined in Lemma 2.2.6.

Table 6 shows the approximated powers of the statistic (2.1) when the x_j's are uniformly spaced. For each case of a sample size m=20 and m=40, one sided significance level 0.025 and 0.5 are considered. A Monte Carlo experiment was performed and shows that the approximation given in Theorem 2.4.3 is accurate enough. The forth column of Table 6 involves the LRS with different choices of δ and ρ. And the sixth column involves the MLRS. Roughly speaking, the unmodified LRS and the modified LRS perform about the same, but it can be seen that the modified LRS with $m_0 > 1$ and $m_1 < m - 1$ has power which improves over the unmodified LRS at points except those close to 0 or m.

To find a confidence region for ρ, the method of Cox and Spijϕtvoll (1982), discussed in Worsely (1986) can be used. Let the confidence region D_α contain all change-points ρ that partition the sequence into two subsequences in which we accept the hypothesis of no further change-points at level α. Consider the tests for a further change-point in two-subsequences;

$H_{o,\rho}^- \; : \; \alpha^{(1)} = \cdots = \alpha^{(\rho)}$ against

$H_{1,\rho}^- \; : \; \exists \; 1 \le k < \rho$ such that $\alpha^{(1)} = \cdots = \alpha^{(k)} \ne \alpha^{(k+1)} = \cdots = \alpha^{(\rho)}$

and

$H_{o,\rho}^+ \; : \; \alpha^{(\rho+1)} = \cdots = \alpha^{(m)}$ against

$$H^+_{1,\rho} \ : \ \exists \ \rho + 1 \le k < m \quad \text{such that} \quad \alpha^{(\rho+1)} = \cdots = \alpha^{(k)} \ne \alpha^{(k+1)} = \cdots = \alpha^{(m)}.$$

If both of $H^-_{0,\rho}$ and $H^+_{0,\rho}$ are accepted at the combined level α, then we put ρ in D_α. Let $M^-_{1,\rho}$ be the equivalent of the test statistic M_1 evaluated only for the subsequence of observations $(x_1, y_1), \ldots, (x_\rho, y_\rho)$ and let $M^+_{1,\rho}$ be the equivalent of the test statistic M_1 evaluated only for the subsequence of observations $(x_{\rho+1}, y_{\rho+1}), \ldots, (x_m, y_m)$ Define $\Pr\{M^-_{1,\rho} < b\} = G^-_\rho(b)$ and $\Pr\{M^+_{1,\rho} < b\} = G^+_\rho(b)$. Then

$$\begin{aligned}\Pr\{M^-_{1,\rho} < b_1 \text{ and } M^+_{1,\rho} < b_2\} &= \Pr\{M^-_{1,\rho} < b_1\} \Pr\{M^+_{1,\rho} < b_2\} \\ &= G^-_\rho(b_1) G^+_\rho(b_2)\end{aligned}$$

and so an exact $(1 - \alpha)$ confidence region for ρ is

$$D_\alpha = \{\rho : G^-_\rho(M^-_{1,\rho}) G^+_\rho(M^+_{1,\rho}) \le 1 - \alpha\}.$$

Asymptotically as $m \to \infty$, and $\rho \to \infty$, $G^-_\rho(\cdot)$ and $G^+_\rho(\cdot)$ has the same formula and can be obtained from (2.16).

In the rest of the section, mathematical results about the confidence set of the change point are stated and the related problems will be discussed. Suppose that we observe $y_1, \ldots, y_m$ and $x_j = j/m$ for $j = 1, \ldots, m$, that the hypotheses of exactly one change only in the intercept of the regression line is true, and that α_1, α_2, and β are unknown nuisance parameters. Then the likelihood based confidence set for a change point can be defined as follows.

For $1 \le m_0 < m_1 < m$ and $c > 0$, define

$$A(\rho, c) = \{\max_{m_0 \le i \le m_1} [U_m(i/m)]^2 - [U_m(\rho/m)]^2 < c\},$$

where U_m is the process defined in Section 2.1.

Although the unconditional probability of $A(\rho, c)$ depends on both ρ and $(\alpha_2 - \alpha_1)$, inference can be made free of $(\alpha_2 - \alpha_1)$ if we condition on the sufficient statistic $U^*_m(\rho) = U_m(\rho/m)\{\rho(1 - \rho/m)\}^{\frac{1}{2}} = \xi$. Thus in principle $c = c(\alpha, \rho, \xi)$ can be determined by

$$\Pr\{A(\rho, c) | U^*_m(\rho) = \xi\} = (1 - \alpha), \tag{2.28}$$

where α is a significance level of the test. Then the set of all ρ such that the sample path $\{U_m(j/m), j = m_0, \ldots, m_1\}$ belongs to $A[\rho, c(\alpha, \rho, U_m^*(\rho))]$ is a $(1-\alpha)100\%$ confidence set for ρ. Then

$$\begin{aligned}\alpha &= \Pr\{\max_{m_0 \le i \le m_1} |U_m(i/m)| \ge b \mid U_m^*(\rho) = \xi\}, \\ &\quad \text{where } b = \{[c(\alpha, \rho, \xi)]^2 + \xi\}^{\frac{1}{2}} \\ &= \Pr_\xi^{(\rho)}\{T_0 < m_1\} \\ &= \Pr_\xi^{(\rho)}\{T_0 < \rho\} + \Pr_\xi^{(\rho)}\{T_1 > \rho\} - \Pr_\xi^{(\rho)}\{T_0 < \rho \text{ and } T_1 > \rho\},\end{aligned} \tag{2.29}$$

where T_0 and T_1 are defined in (2.22). Since the third conditional probability in (2.29) is negligible comparing to the first and the second probabilities which are usually small, in order to get a confidence set it suffices to find an approximation to $\Pr_\xi^{(\rho)}\{T_0 < \rho\}$. This conditional probability depends on how big the difference between the conditioned value, ξ, and the boundary value at the change point, $\pm b\{\rho(1-\rho/m)\}^{\frac{1}{2}}$. In this section, we consider the confidence set when

$$\Delta = b\{\rho(1-\rho/m)\}^{\frac{1}{2}} - \xi = O(m).$$

Then

$$\begin{aligned}P &= \Pr_\xi^{(\rho)}\{T_0 < \rho\} \\ &= \sum_{n=m_0}^{\rho-1} \Pr_\xi^{(\rho)}\{T_0 = n\} \\ &= \sum_{n=m_0}^{\rho-1} \int_0^\infty \Pr\{|U_{m,\xi}^{*,(\rho)}(k)| \le b\{k(1-k/m)\}^{\frac{1}{2}}, \text{for all } m_0 \le k < n \mid |U_{m,\xi}^{*,(\rho)}(n)| = y\} \\ &\qquad \times \Pr_\xi^{(\rho)}\{|U_m^*(n)| \in b\{n(1-n/m)\}^{\frac{1}{2}} + dx\},\end{aligned}$$

where $U_{m,\xi}^{*,(\rho)}(k)$ is the process $U_{m,\xi}^*(k)$ conditioned on $U_m^*(\rho) = \xi$, and $y = b\{n(1-n/m)\}^{\frac{1}{2}} + x$.

Lemma 2.4.4.

Given that $U_{m,\xi}^{*,(\rho)}(n) = y$, as $m \to \infty, n/m \to s, \rho/m \to t, \xi = m\xi_0$, and for a fixed k, $[U_{m,\xi}^{*,(\rho)}(n-k) - \mu_\xi^{(\rho)}(n-k \mid n)]\{D(s,s)\}^{\frac{1}{2}}$ is distributed approximately as sum of *i.i.d.* random variables each of which has mean 0 and variance 1, where

$$
\begin{aligned}
\mu_\xi^{(\rho)}(k|n) &= k[C_1(k,n,\rho)\xi_0 + C_2(k,n,\rho)y/n], \\
C_1(k,n,\rho) &= \frac{(1-n/m)[r(k,\rho) - r(k,n)r(n,\rho)]}{[1-(n/m) - r^2(n,\rho)(1-\rho/m)n/\rho]\rho/m}, \\
C_2(k,n,\rho) &= \frac{(1-n/m)r(k,n) - r(k,\rho)r(n,\rho)(1-\rho/m)n/\rho}{[1-(n/m) - r^2(n,\rho)(1-\rho/m)n/\rho]}, \\
r(n,\rho) &= D(n/m,\rho/m)/\{D(n/m,n/m)D(\rho/m,\rho/m)\}^{\frac{1}{2}} \\
D(n/m,\rho/m) &= 1 - 3(1-n/m)(\rho/m) \ \text{ for } n < \rho.
\end{aligned}
$$

Proof. Since $U_{m,\xi}^{*,(\rho)}(n)$ is distributed as $N(\mu(n|\rho), \sigma^2(n|\rho))$, where

$$
\begin{aligned}
\mu(n|\rho) &= \xi r_m(n,\rho)n/\rho, \\
\sigma^2(n|\rho) &= n(1-n/m) - \{r_m(n,\rho)\}^2(1-\rho/m)n^2/\rho.
\end{aligned}
$$

it can be shown that

$$
\begin{aligned}
\mu_\xi^{(\rho)}(k|n) &= E\,[\,U_{m,\xi}^{*,(\rho)}(k)\mid U_\xi^{*,(\rho)}(n) = y\,] \\
&= k[C_1(k,n,\rho)\xi_0 + C_2(k,n,\rho)y/n], \\
[\sigma_\xi^{(\rho)}(k|n)]^2 &= \mathrm{Var}\,[\,U_{m,\xi}^{*,(\rho)}(k)\mid U_{m,\xi}^{*,(\rho)}(n) = y\,] \\
&= k[\zeta(k,\rho) - \{C_1(k,n,\rho)\}^2\zeta(n,\rho)k/n],
\end{aligned}
$$

where

$$\zeta(k,\rho) = 1 - (k/m) - r_m^2(k,\rho)(1-\rho/m)k/\rho.$$

Then direct calculation implies that as $m \to \infty$,

$$[\sigma_\xi^{(\rho)}(n-k\mid n)]^2 \to k/D(s,s),$$

and

$$\mathrm{Cov}[U_{m,\xi}^{*,(\rho)}(n-k_1), U_{m,\xi}^{*,(\rho)}(n-k_2)\mid U_{m,\xi}^{*,(\rho)}(n) = y] \to \min(k_1,k_2)/D(s,s).$$

Hence the proof is completed. ∎

Lemma 2.4.5.

Suppose that $b \to \infty, m \to \infty$, in such a way that $n/m \to s$, $\rho/m \to t$, and $b/\sqrt{m} \to b_0$. Then

$$\begin{aligned} P_b(n,x) &= \Pr\{U^{*,(\rho)}_{m,\xi}(k) \le b\{k(1-k/m)^{\frac{1}{2}}\} \quad \text{for all } m_0 \le k < n \mid U^{*,(\rho)}_{m,\xi}(n) = y\} \\ &\cong \Pr{}_{\mu}\{S_k \ge x\{D(s,s)\}^{\frac{1}{2}}, \ \text{for all } k \ge 1\}, \end{aligned}$$

where S_k is the sum of k *i.i.d.* random variables with mean μ and variance 1, and

$$\begin{aligned} \mu &= B_1 - B_2\xi_0, \\ B_1 &= b_0\{D(s,s)\}^{\frac{1}{2}}/[2\{s(1-s)\}^{\frac{1}{2}}], \\ B_2 &= 3(1-s)(s/t-1)n/[m\zeta(n,\rho)D(s,s)\{D(t,t)\}^{\frac{1}{2}}], \\ D(t_1,t_2) &= 1 - 3t_2(1-t_1) \ \text{ for } t_1 < t_2. \end{aligned}$$

Proof. Note that

$$P_b(n,x) = \Pr\{U^{*,(\rho)}_{m,\xi}(k) - \mu^{(\rho)}_{\xi}(k|n) \le B_b(n,k), \text{ for all } \ m_0 \le k < n \mid U^{*,(\rho)}_{m,\xi}(n) = y\},$$

where

$$B_b(n,k) = b\{k(1-k/m)\}^{\frac{1}{2}} - \mu^{(\rho)}_{\xi}(k|n).$$

Since the joint distribution of $\{[U^{*,(\rho)}_{m,\xi}(n-k) - \mu^{(\rho)}_{\xi}(n-k \mid n)]\{D(n/m,n/m)\}^{\frac{1}{2}}, k = 1,\ldots,n-m_0\}$ given that $U^{*,(\rho)}_{m,\xi}(n) = y$ converges to the unconditional joint distribution of $\{S_k, k = 1,\ldots,n-m_0\}$ and $B_b(n,n-k)\{D(n/m,n/m)\}^{\frac{1}{2}} \sim k[B_1 - B_2\xi_0] - x\{D(s,s)\}^{\frac{1}{2}}$,

$$\begin{aligned} P_b(n,x) &\cong \Pr{}_0\{S_k \le k[B_1 - B_2\xi_0] - x\{D(s,s)\}^{\frac{1}{2}}, \text{ for all } k \ge 1\} \\ &= \Pr{}_{-\mu}\{S_k \le -x\{D(s,s)\}^{\frac{1}{2}}, \text{ for all } k \ge 1\} \\ &= \Pr{}_{\mu}\{S_k \ge x\{D(s,s)\}^{\frac{1}{2}}, \text{ for all } k \ge 1\} \\ &= \{E[S_{\tau_+}]\}^{-1}\Pr{}_{\mu}\{S_{\tau_+} \ge x\{D(s,s)\}^{\frac{1}{2}}\}\mu, \end{aligned}$$

where $\tau_+ = \inf\{k : k \ge 1, \ S_k > 0\}$ and the last equality holds by the argument in Siegmund (1986, Chapter XIII). ∎

Theorem 2.4.6.

Assume that $b\{\rho_1 - \rho/m)\}^{\frac{1}{2}} - \xi = O(m)$ and $b \to \infty$, $n \to \infty$, $\rho \to \infty$, and $m \to \infty$, in such a way that $b/\sqrt{m} \to b_0$, $n/m \to s$, and$\rho/m \to t$. Then

$$\Pr{}^{(\rho)}_{\xi}\{T < \rho\}$$
$$\cong \sum_{n=m_0}^{\rho-1} \int_0^\infty \exp[-d(n,\rho)x/\sigma(n,\rho)] \Pr{}_\mu\{S_{\tau_+} \geq x\{D(s,s)\}^{\frac{1}{2}}\} dx R(n,\rho),$$

where

$$R(n,\rho) = \phi(d(n,\rho))\mu\{E_\mu[S_{\tau_+}]\}\}^{-1},$$
$$d(n,\rho) = [b\{n(1-n/m)\}^{\frac{1}{2}} - \mu(n|\rho))]/\sigma(n|\rho).$$

Proof. Note that

$$\Pr{}^{(\rho)}_{\xi}\{T_0 < \rho\}$$
$$= \Pr{}^{(\rho)}_{\xi}\{\tau_0^+ < \rho\} + \Pr{}^{(\rho)}_{\xi}\{\tau_0^- < \rho\} - \Pr{}^{(\rho)}_{\xi}\{\tau_0^+ < \rho \text{ and } \tau_0^- < \rho\}.$$

Since for $\xi > 0$,

$$\Pr{}^{(\rho)}_{\xi}\{T_0 < \rho\} \cong \Pr{}^{(\rho)}_{\xi}\{\tau_0^+ < \rho\}$$
$$= \sum_{n=m_0}^{\rho-1} \Pr{}^{(\rho)}_{\xi}\{\tau_0^+ = n\}$$
$$\cong \sum_{n=m_0}^{\rho-1} \int_0^\infty P_b(n,x) \Pr{}^{(\rho)}_{\xi}\{U_m^*(n) \in b\{n(1-n/m)\}^{\frac{1}{2}} + dx\},$$

and similarly

$$\Pr{}^{(\rho)}_{-\xi}\{T_0 < \rho\} \cong \Pr{}^{(\rho)}_{-\xi}\{\tau_0^- < \rho\},$$

by Lemmas 2.4.4. and 2.4.5 and letting $m \to \infty$,

$$\Pr{}^{(\rho)}_{\xi}\{T_0 < \rho\}$$
$$\cong \sum_{n=m_0}^{\rho-1} \int_0^\infty \{\Pr{}_\mu\{S_{\tau_+} \geq x\{D(s,s)\}^{\frac{1}{2}}\}/E_\mu[S_{\tau_+}]\} dx \{\phi(d(n,\rho)+x)/\sigma(n|\rho)\}$$
$$\cong \sum_{n=m_0}^{\rho-1} \int_0^\infty \Pr{}_\mu\{S_{\tau_+} \geq x\{D(s,s)\}\} \exp[-d(n,\rho)x/\sigma(n|\rho)] dx R(n,\rho),$$

where $d(n,\rho)$ and $R(n,\rho)$ are defined above. ∎

Remark 2.5. If $d(n^*, \rho)/[\sigma(n^*|\rho)\{D(n^*/m, n^*/m)\}^{\frac{1}{2}}] = 2\mu$ for some n^* at which the integration has the biggest contribution to $\Pr_\xi^{(\rho)}\{T_0 < \rho\}$, then $\Pr_\xi^{(\rho)}\{T_0 < \rho\}$ can be reduced to $\sum \Pr_\xi^{(\rho)}\{T_0 = n\}$, where the summation is over n which are close to n^*, and to a further simpler form by the argument in Siegmund (1986,Chapter IX).

Chapter 3

Change in Both Intercept and Slope

3.1. Models and Test Statistics

In Chapter 3, the two-phase linear regression model in which both intercept and slope terms change will be considered. Quandt (1958) introduced this model. He proposed the LRT to test for this type of two-phase regression model as opposed to the null hypothesis of the simple linear regression and observed that the LRS doesn't follow the standard maximum likelihood asymptotic theory. This type of two-phase regression model has many applications in econometrics, biology, quality control, and so on. Brown, Durbin and Evans (1975) give three examples involving growth in the number of local telephone calls, the demand for money, and staff requirements of an organization. They use recursive residuals to study the stability over time of regression relationships and discuss Quandt's likelihood method. Hinkely (1971) studies a small set of data obtained from replicated experimental determination of the relationship between blood factor VII production and wafarin concentration. He applies a broken line regression model with a continuity constraint to this set of data. The same kind of example appears in Haddad, Jeng, and Lai (1987) who use a two-phase regression model to summarize the time course and change in heart rate during respiratory pauses in puppies and young adult dogs.

We consider the problem of testing the null hypothesis that the data follow one simple linear regression :

$$H_0 \; : \; y_j = \alpha + \beta x_j + \varepsilon_j, \qquad j = 1, \ldots, m, \qquad \text{against}$$

the alternative hypothesis that there is a change both in the intercept and slope :

$$
\begin{aligned}
H_1 \; : \; \exists \; 1 \le \rho < m \quad & \text{such that} \\
& y_j = \alpha_1 + \beta_1 x_j + \varepsilon_j, \qquad j = 1, \ldots, \rho \\
& y_j = \alpha_2 + \beta_2 x_j + \varepsilon_j, \qquad j = \rho + 1, \ldots, m,
\end{aligned}
$$

where $\alpha_1 \neq \alpha_2$ and$\beta_1 \neq \beta_2$.

Unlike Hinkely's model, we do not assume mathematical continuity of the two-phase regression line and we suppose that a change happens at the ρ^{th} data point as in Chapter 2. In this section, we assume that the ε_j's are independently, identically normally distributed with mean 0 and variance σ^2 and we derive the LRS for cases of known and unknown σ^2 and study the null and alternative distributions of the LRS. When σ^2 is known, σ^2 can be assumed to be equal to 1 without loss of generality. Then -2log(likelihood ratio) statistic for a fixed change point i is proportional to

$$[\bar{y}_m - \bar{y}_i]^2 mi/(m-i) + [Q_{xy,i}^2/Q_{xx,i} + Q_{xy,i}^{*\,2}/Q_{xx,i}^*] - [Q_{xy,m}^{*\,2}/Q_{xx,m}], \qquad (3.1)$$

where

$$
\begin{aligned}
\bar{x}_i &= \Big(\sum_{j=1}^{i} x_j\Big)/i, & \bar{x}_i^* &= \Big(\sum_{j=i+1}^{m} x_j\Big)/(m-i), \\
\bar{y}_i &= \Big(\sum_{j=1}^{i} y_j\Big)/i, & \bar{y}_i^* &= \Big(\sum_{j=i+1}^{m} y_j\Big)/(m-i), \\
Q_{xx,i} &= \sum_{j=1}^{i} (x_j - \bar{x}_i)^2, & Q_{xx,i}^* &= \sum_{j=i+1}^{m} (x_j - \bar{x}_i^*)^2, \\
Q_{xy,i} &= \sum_{j=1}^{i} (x_j - \bar{x}_i)(y_j - \bar{y}_i), & Q_{xy,i}^* &= \sum_{j=i+1}^{m} (x_j - \bar{x}_i^*)(y_j - \bar{y}_i^*).
\end{aligned}
$$

To get some insight about the distribution of (3.1), we can rewrite (3.1) as

$$\| \mathbf{V}_m(i/m) \|^2 = \hat{\delta}_i' \Sigma_i^{-1} \hat{\delta}_i,$$

where

$$\hat{\delta}_i = (\hat{\alpha}_i - \hat{\alpha}_i^*, \hat{\beta}_i - \hat{\beta}_i^*),$$

$$\hat{\alpha}_i = \bar{y}_i - \hat{\beta}_i\bar{x}_i, \qquad \hat{\alpha}_i^* = \bar{y}_i^* - \hat{\beta}_i^*\bar{x}_i^*,$$

$$\hat{\beta}_i = Q_{xy,i}/Q_{xx,i}, \qquad \hat{\beta}_i^* = Q_{xy,i}^*/Q_{xx,i}^*,$$

$$\Sigma = \begin{pmatrix} m/[i(m-i)] + (\bar{x}_i^2/Q_{xx,i}) + (\bar{x}_i^{*2}/Q_{xx,i}^*) & \bar{x}_i/Q_{xx,i} + \bar{x}_i^*/Q_{xx,i}^* \\ \bar{x}_i/Q_{xx,i} + \bar{x}_i^*/Q_{xx,i}^* & 1/Q_{xx,i} + 1/Q_{xx,i}^* \end{pmatrix}.$$

Hence the likelihood ratio test (LRT) of H_0 against H_1 can be based on

$$\max_{1 \le i < m} \| \mathbf{V}_m(i/m) \| .$$

As in Chapter 2, we shall consider the modified LRS

$$M_3 = \max_{m_0 \le i \le m_1} \| \mathbf{V}_m(i/m) \| , \tag{3.2}$$

where $1 < m_0 < m_1 < m$. Based on the MLRS, H_0 is rejected for a large value of M_3 and the value of i which maximizes $\| \mathbf{V}_m(i/m) \|$ is the maximum likelihood estimate of the true change point. Under H_0, $\hat{\delta}_i$ has a bivariate normal distribution with mean (0,0) and covariance matrix Σ_i for each i and so the null distribution of M_3 is the maximum of a sequence of correlated chi-square random variables.

Here is another expression of the LRS which will be used in Sections 3.2 and 3.3 :

$$\| \mathbf{V}_m(i/m) \|^2 = [V_{1,m}(i/m)]^2 + [V_{2,m}(i/m)]^2, \tag{3.3}$$

where

$$V_{1,m}(i/m) = [A_1'(X_1^{i\prime}X_1^i)^{-1}X_1^{i\prime}Y]/[A_1'(X_1^{i\prime}X_1^i)^{-1}A_1]^{\frac{1}{2}},$$

$$V_{2,m}(i/m) = [A_2'(X_2^{i\prime}X_2^i)^{-1}X_2^{i\prime}Y]/[A_2'(X_2^{i\prime}X_2^i)^{-1}A_2]^{\frac{1}{2}},$$

$$A_1' = (1, -1, 0), \qquad A_2' = (0, 1, 0, -1),$$

$$X_1^i = \begin{pmatrix} 1 & 0 & x_1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & x_i \\ 0 & 1 & x_{i+1} \\ \vdots & \vdots & \vdots \\ 0 & 1 & x_m \end{pmatrix}, \qquad X_2^i = \begin{pmatrix} 1 & x_1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_i & 0 & 0 \\ 0 & 0 & 1 & x_{i+1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & x_m \end{pmatrix},$$

$$Y' = (y_1, \ldots, y_m).$$

From this representation, one sees that $V_{1,m}(i/m)$ is the same statistic as $U_m(i/m)$ defined in Section 2.1. That is, $V_{1,m}(i/m)$ is the LRS to test that only the intercept changes as opposed to the null hypothesis of no change. On the other hand $V_{2,m}(i/m)$ is the LRS to test the null hypothesis that only the intercept term in the regression line changes at the point i against the alternative hypothesis that both intercept and slope change after the same point i. It is easy to show that each of $[V_{1,m}(i/m)]^2$ and $[V_{2,m}(i/m)]^2$ has a chi-square distribution with 1 degree of freedom and the covariance function of the process $\{V_{1,m}(i/m),\ i = 1, \ldots, m\}$ was given in Section 2.1. For the process $V_{2,m}$ the covariance between $V_{2,m}(i/m)$ and $V_{2,m}(j/m)$ for $i < j$ is given by

$$\mathrm{Cov}\,[\, V_{2,m}(i/m), V_{2,m}(j/m)\,] = \left\{ \frac{Q_{xx,i} Q^*_{xx,j}}{Q^*_{xx,i} Q_{xx,j}} \right\}^{\frac{1}{2}} \frac{D_m(i/m, j/m)}{\{D_m(i/m, i/m) D_m(j/m, j/m)\}^{\frac{1}{2}}}, \quad (3.4)$$

where

$$D_m(i/m, j/m) = 1 - (\bar{x}_m - \bar{x}_i)(\bar{x}_m - \bar{x}_j) mj/[(m-j)Q_{xx}] \qquad \textit{for} \quad i < j.$$

It is convenient to introduce the notations

$$\lambda_{11} = \mathrm{Cov}\,[V_{1,m}(i/m), V_{1,m}(j/m)], \quad \lambda_{12} = \mathrm{Cov}\,[V_{1,m}(i/m), V_{2,m}(j/m)]$$

$$\lambda_{21} = \mathrm{Cov}\,[V_{1,m}(j/m), V_{2,m}(i/m)], \quad \lambda_{22} = \mathrm{Cov}\,[V_{2,m}(i/m), V_{2,m}(j/m)].$$

One delicate matter is the cross covariance between the two processes $V_{1,m}$ and $V_{2,m}$. It can be easily checked that $V_{1,m}(i/m)$ and $V_{2,m}(i/m)$ are independent at each point i. However for different points i and j such that $i < j$ covariance function is as follows:

$$\lambda_{12} = \left\{ \frac{miQ^*_{xx,j}}{(m-i)Q_{xx,m}Q_{xx,j}} \right\}^{\frac{1}{2}} \frac{(\bar{x}_i - \bar{x}_j)}{\{D_m(i/m, i/m)D_m(j/m, j/m)\}^{\frac{1}{2}}}, \tag{3.5}$$

$$\lambda_{21} = \left\{ \frac{m(m-j)Q_{xx,i}}{jQ_{xx,m}Q^*_{xx,i}} \right\}^{\frac{1}{2}} \frac{(\bar{x}^*_j - \bar{x}^*_i)}{\{D_m(i/m, i/m)D_m(j/m, j/m)\}^{\frac{1}{2}}}. \tag{3.6}$$

In summary,

$$\mathrm{Cov}\,[\, \mathbf{V}_m(i/m), \mathbf{V}_m(j/m)\,] = \begin{pmatrix} I_2 & \Lambda \\ \Lambda & I_2 \end{pmatrix},$$

where

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix}.$$

Thus $\{\mathbf{V}_m(i/m), i = 1, \ldots, m\}$ is the two dimensional stochastic process with zero drift and the covariance function given above. Again the null distribution of M_3 depends on the x_j's only through this covariance structure of $\{\mathbf{V}_m(i/m)\}$, not on α, β. Under the alternative, $\mathbf{V}_m(i/m)$ has a bivariate normal distribution and the covariance structure remains the same as under the null hypothesis. So the only difference of the LRS under H_1 is non-zero drift of $\{\mathbf{V}_m(i/m)\}$. For convenience we use the notation

$$\Delta_\alpha = \alpha_2 - \alpha_1, \qquad \Delta_\beta = \beta_2 - \beta_1.$$

Then under H_1, $V_{1,m}(i/m)$ has non-zero mean for all i, which is

$$\begin{aligned} &E\,[V_{1,m}(i/m)] \\ &\quad = i\, \frac{\left[(1-\rho/m)A_{i,\rho}\{\Delta_\alpha + \Delta_\beta \bar{x}^*_\rho\} - \Delta_\beta(\bar{x}_i - \bar{x})Q^*_{xx,\rho}\right]}{\{i(1-i/m)D_m(i/m, i/m)Q_{xx,m}\}^{\frac{1}{2}}}, \qquad i \le \rho \\ &\quad = (m-i)\, \frac{\left[(\rho/m)A_{i,\rho}\{\Delta_\alpha + \Delta_\beta \bar{x}_\rho\} - \Delta_\beta(\bar{x}^*_i - \bar{x})Q_{xx,\rho}\right]}{\{i(1-i/m)D_m(i/m, i/m)Q_{xx,m}\}^{\frac{1}{2}}}, \quad i > \rho, \end{aligned}$$

where

$$A_{i,\rho} = -Q_{xx,m}D_m(i/m, \rho/m).$$

And for $V_{2,m}(i/m)$,

$$
\begin{aligned}
E\,[V_{2,m}(i/m)]& \\
&= \left\{\frac{Q_{xx,i}}{Q^{*}_{xx,i}}\right\}^{\frac{1}{2}} \frac{\left[(m-\rho)(\bar{x}^{*}_{i}-\bar{x}^{*}_{\rho})\{\Delta_{\alpha}+\Delta_{\beta}\bar{x}^{*}_{\rho}\}-\Delta_{\beta}Q^{*}_{xx,\rho}\right]}{\{D_m(i/m,i/m)\}^{\frac{1}{2}}}, \qquad i \le \rho \\
&= \left\{\frac{Q^{*}_{xx,i}}{Q_{xx,i}}\right\}^{\frac{1}{2}} \frac{\left[\rho(\bar{x}_{i}-\bar{x}_{\rho})\{\Delta_{\alpha}+\Delta_{\beta}\bar{x}_{\rho}\}-\Delta_{\beta}Q_{xx,\rho}\right]}{\{D_m(i/m,i/m)\}^{\frac{1}{2}}}, \qquad i > \rho.
\end{aligned}
$$

So the alternative distribution of M_3 depends on unknown parameters $\alpha_2-\alpha_1$, $\beta_2-\beta_1$ and the unknown change point ρ.

If σ^2 is unknown, the LRS is proportional to

$$\max_{1\le i<m} \| \mathbf{V}_m(i/m) \| / \hat{\sigma}$$

where $\hat{\sigma}^2 = (Q_{yy,m} - Q^2_{xy,m}/Q_{xx,m})/m$. Thus the modified LRS is

$$M_4 = \max_{m_0\le i\le m_1} \| \mathbf{V}_m(i/m) \| \;/\; \hat{\sigma}.$$

In the following sections, similar kinds of results as in Chapter 2 will be discussed. We study the asymptotic behavior of the MLRS under H_0 for the cases of known and unknown variance. In Section 3.3, we derive an approximation to the significance level of M_3 and present simulation results which support the analytical approximation derived for known variance case, and show that this approximation can be applied for unknown variance case.

3.2. Asymptotic Behavior of Test Statistics

In Chapter 2, it was seen that the MLRS converges to the maximum absolute value of functions of Brownian bridge processes or Gaussian processes according to the random or fixed x_j's, respectively. In the case where both intercept and slope change, we shall obtain similar results which are extensions of those of Section 2.2. As we can guess from the form of the MLRS, the limiting distribution is the maximum norm of random functions involving two-dimensional Brownian bridges or two-dimensional Gaussian processes.

Section 3.2.1 concerns the case in which the independent variable is random. Also, the asymptotic behavior of the MLRS is considered conditionally on the x_j's. As in Chapter 2 we obtain the same limiting distributions whether we consider the null distribution of the MLRS conditionally or unconditionally. We will deal with the case of the fixed values of the independent variable in Section 3.2.3. The limiting behavior of the MLRS under a mild assumption about the values of the independent variable will be studied, starting from the case where the values of the independent variable are uniformly spaced. Although the MLRS was derived assuming that the ϵ_j's are identically and normally distributed, the asymptotic results to be discussed in Sections 3.2.1 and 3.2.2 do not require this normality umption.

3.2.1. When the independent variable is random

This section will show similar results as in Section 2.2.1 using Donsker's theorem when the independent variable is also random. \s in Section 2.2.1, it can be easily checked that the MLRS does not depend on the slope under the null hypothesis. This implies that we can take x as the random variable which is independent of y when we study the null distribution of the MLRS. The following theorem is on the convergence in distribution of the MLRS when σ_x^2 and σ_y^2 are known. In this case we may assume $\sigma_x^2 = \sigma_y^2 = 1$ without loss of generality.

Theorem 3.2.1.

Let $(x_1, y_1), \ldots, (x_m, y_m)$ be i.i.d. random variables such that $E[x_1] = E[y_1] = 0$, $E[{x_1}^2] = E[{y_1}^2] = 1$, and $E[x_1 y_1] = 0$.

Under H_0, as $m \to \infty$ and $m_i/m \to t_i$ for $i = 0, 1$,

$$M_3 = \max_{m_0 \le i \le m_1} \frac{\| \mathbf{V}_m^0(i/m) \|}{\{(i/m)(1 - i/m)\}^{\frac{1}{2}}} \to \max_{t_0 \le t \le t_1} \frac{\| \mathbf{W}^0(t) \|}{\{t(1-t)\}^{\frac{1}{2}}} \quad \text{in distribution.}$$

where $\mathbf{V}_m^0(i/m) = \mathbf{V}_m(i/m)\{(i/m)(1 - i/m)\}^{\frac{1}{2}}$, $\mathbf{W}^0(t)' = (W_1^0(t), W_2^0(t))$, W_1^0 and W_2^0 are two independent Brownian bridge processes.

Proof : Note that

$$\|\mathbf{V}_m^0(i/m)\|^2 = [V_{1,m}^0(i/m)]^2 + [V_{2,m}^0(i/m)]^2,$$

where

$$V_{1,m}^0(i/m) = [B_y(i/m) - B_x(i/m)(Q_{xy,m}/Q_{xx,m})]/\{D_m(i/m, i/m)\}^{\frac{1}{2}},$$

$$V_{2,m}^0(i/m) = \{1/Q_{xx,i} + 1/Q^*_{xx,i}\}^{-\frac{1}{2}}\{Q_{xy,i}/Q_{xx,i} - Q^*_{xy,i}/Q^*_{xx,i}\},$$

$$D_m(i/m, i/m) = 1 - [B_x(i/m)]^2/[(i/m)(1 - i/m)Q_{xx,m}],$$

$$B_x(i/m) = (\bar{x}_i - \bar{x}_m)i/\sqrt{m}, \qquad B_y(i/m) = (\bar{y}_i - \bar{y}_m)i/\sqrt{m}.$$

(i) Let

$$Y_m(i/m) = B_y(i/m)/\{D_m(i/m, i/m)\}^{\frac{1}{2}},$$

$$X_m(i/m) = B_x(i/m)(Q_{xy,m}/Q_{xx,m})/\{D_m(i/m, i/m)\}^{\frac{1}{2}},$$

so that $V_{1,m}^0(i/m) = Y_m(i/m) - X_m(i/m)$. In Theorem 2.2.2, we have shown that

$$Y_m \to W_1^0 \quad \text{in distribution and} \quad X_m \to 0 \quad \text{in probability},$$

which leads to

$$V_{1,m}^0 \to W_1^0 \qquad \text{in distribution.}$$

(ii) Note that $V_{2,m}^0(i/m)$ can be rewritten as

$$\left\{\frac{(Q_{xx,i}/i)(Q^*_{xx,i}/(m-i))}{(Q_{xx,m}/m) - [Z_x(i/m)]^2/(i(1-i/m))}\right\}^{\frac{1}{2}} \times \left(\{1 - i/m\}^{\frac{1}{2}}\frac{Q_{xy,i}/\sqrt{i}}{Q_{xx,i}/i} - \{i/m\}^{\frac{1}{2}}\frac{Q^*_{xy,i}/\sqrt{m-i}}{Q^*_{xx,i}/(m-i)}\right).$$

Since $V_{2,m}^0(i/m)$ is a function of the partial sum, $\sum_{j=1}^{i} x_j y_j$, Donsker's Theorem can be used to show that $V_{2,m}^0 \to W_2^0$ in distribution in the following way. Let $W_m(i/m) = Q_{xy,i}/\sqrt{m}$. First we use the convergence of W_m to the Brownian motion W_2 to show that for any sets of $(r_1, \ldots, r_n)$ and $(i_1, \ldots, i_n)$ such that $(i_1/m, \ldots, i_n/m) \to (t_1, \ldots, t_n)$ as $m \to \infty$,

$$E[\exp\{i\sum_{k=1}^{n} r_k V_{2,m}^0(i_k/m)\}] = E[\exp\{i\sum_{k=1}^{n}(c_{i_k}Q_{xy,i_k} + c^*_{i_k}Q^*_{xy,i_k})\}]$$

$$\to E[\exp\{i\sum_{k=1}^{n} r_k W_2^0(t_k)\}] \qquad \text{in distribution,}$$

where c_{i_k} and $c^*_{i_k}$ are appropriate coefficients. This implies that the finite dimensional distributions of $V^0_{2,m}$ converge to those of W^0_2. Secondly, the tightness of the sequence $\{V^0_{2,m}\}$ follows from the same sort of argument involved in the proof of Donsker's theorem (Billingsley,1968). Hence

$$V^0_{2,m} \to W^0_2 \qquad \text{in distribution.}$$

(iii) Since $V^0_{2,m}$ and Y_m are independent and $X_m \to 0$ in probability, it is easy to show that $(X_m, Y_m, V^0_{2,m}) \to (0, W^0_1, W^0_2)$ in distribution.

Then by the continuous mapping theorem the proof is completed. ∎

As pointed out in Section 2.2, we obtain the same limiting distribution as in the preceding theorem when the variances are unknown.

Corollary 3.2.2.

Under the same assumptions as in Theorem 3.2.1,

$$M_4 = \max_{m_0 \le i \le m_1} \frac{\| \tilde{\mathbf{V}}^0_m(i/m) \|}{\{(i/m)(1-i/m)\}^{\frac{1}{2}}} \to \max_{t_0 \le t \le t_1} \frac{\| \mathbf{W}^0(t) \|}{\{t(1-t)\}^{\frac{1}{2}}} \quad \text{in distribution,}$$

where $\tilde{\mathbf{V}}^0_m(i/m) = \tilde{\mathbf{V}}^0_m(i/m)\{(i/m)(1-i/m)\}^{\frac{1}{2}}$.

In the following theorem, the asymptotic behavior of the MLRS will be considered conditionally on the x_j's when the x_j's are a random sample from some distribution.

Theorem 3.2.3.

Let $\mathbf{z}'_j = (x_j, y_j),\ j = 1, \ldots, m$ be a sequence of i.i.d. random vectors such that $E[\mathbf{z}_j] = \boldsymbol{\mu}$ and $E[\mathbf{z}_j\mathbf{z}'_j] = \Sigma = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}$.

Under H_0, as $m \to \infty$ and $m_i/m \to t_i$ for $i = 0, 1$, conditionally given $x_1, x_2, \ldots,$

$$M_3 = \max_{m_0 \le i \le m_1} \frac{\|\mathbf{V}^0_m(i/m)\|}{\{(i/m)(1-i/m)\}^{\frac{1}{2}}} \to \max_{t_0 \le t \le t_1} \frac{\|\mathbf{W}^0(t)\|}{\{t(1-t)\}^{\frac{1}{2}}} \quad \text{in distribution.}$$

for a.e. $x_1, x_2, \cdots$, where $\mathbf{W}^0(t)' = (W_1^0(t), W_2^0(t))$ and W_1^0 and W_2^0 are two independent Brownian bridges.

Proof : To prove this, we follow the same argument as in the proof of Theorem 3.2.1. In the proof of Theorem2.2.4, we wrote $V_{1,m}^0 = Z_m - R_m$, where Z_m and R_m were defined in Theorem 2.2.4, and showed that a.e. in x

$$\text{(i) } Z_m \to W_1^0 \qquad \text{in distribution,}$$

$$\text{(ii) } \max_{m_0 \leq i \leq m_1} R_m(i/m) \to 0 \qquad \text{in probability.}$$

Then similar arguments show that, a.e. in x, as $m \to \infty$,

$$V_{2,m}^0 \to W_2^0 \qquad \text{in distribution,}$$

and hence

$$(Z_m, \max_{m_0 \leq i \leq m_1} R_m(i/m), V_{2,m}^0) \to (W_1^0, 0, W_2^0) \qquad \text{in distribution.}$$

Therefore by the continuous mapping theorem, the proof is completed. The independence between W_1^0 and W_2^0 can be proved examining the limiting behavior of the covariance functions given in (3.5) and (3.6). ∎

Corollary 3.2.4.

Under the same assumptions as in Theorem 3.2.3, conditionally given $x_1, x_2, \ldots$,

$$M_4 = \max_{m_0 \leq i \leq m_1} \frac{\| \tilde{\mathbf{V}}_m^0(i/m) \|}{\{(i/m)(1-i/m)\}^{\frac{1}{2}}} \to \max_{t_0 \leq t \leq t_1} \frac{\| \mathbf{W}^0(t) \|}{\{t(1-t)\}^{\frac{1}{2}}} \quad \text{in distribution,}$$

for a.e. $x_1, x_2, \ldots,$ where $\tilde{\mathbf{V}}_m^0(i/m) = \mathbf{V}_m^0(i/m)\{(i/m)(1-i/m)\}^{\frac{1}{2}}$.

In the above Theorem 3.3.3, W_1^0 and W_2^0 are independent since $\bar{x}_i - \bar{x}_j \to 0$ and $\bar{x}_j^* - \bar{x}_i^* \to 0$ as $i, j \to \infty$. Thus the MLRS may not converge to the limiting distribution given above if the empirical distribution of the x_j's does not satisfy these conditions. We will discuss this more carefully in the following section describing how covariance structure depends on the spacings of the x_j's.

3.2.2. When the independent variable is fixed

This section will show that the MLRS, M_3 and M_4 converge to the maximum norm of two-dimensional Gaussian processes when the x_j's are fixed. In the preceding section, it was seen that V_1 and V_2 are asymptotically independent conditionally on the x_j's when the first and second sample moments of the x_j's converge. The covariance function which was given in (3.4)-(3.6) explains the effect of the spacing of the x_j's on the distribution of the MLRS. In the following theorem that gives the limiting distribution of the MLRS, we use the representation of $(V^0_{1,m}(i/m), V^0_{2,m}(i/m))$ as

$$(\sum_{k=1}^{m} a_{i,k}\epsilon_k, \sum_{k=1}^{m} b_{i,k}\epsilon_k),$$

where $a_{i,k}$ was given in (2.6) and

$$\begin{aligned} b_{i,k} &= \left\{\frac{(i/m)(1-i/m)Q_{xx,i}Q^*_{xx,i}}{D_m(i/m,i/m)Q_{xx,m}}\right\}^{\frac{1}{2}} \frac{x_k - \bar{x}_i}{Q_{xx,i}}, \qquad k \le i \\ &= \left\{\frac{(i/m)(1-i/m)Q_{xx,i}Q^*_{xx,i}}{D_m(i/m,i/m)Q_{xx,m}}\right\}^{\frac{1}{2}} \frac{x_k - \bar{x}^*_i}{Q^*_{xx,i}}, \qquad k > i. \end{aligned}$$

Here we assume that σ^2 is known and hence without loss of generality equals one and begin with the case in which the x_j's are uniformly spaced.

Theorem 3.2.5.

Suppose that $x_j = j/m$ for $j = 1, \ldots, m$.

Under H_0, as $m \to \infty$ and $m_i/m \to t_i$ for $i = 0, 1$,

$$M_3 = \max_{m_0 \le i \le m_1} \| \mathbf{V}_m(i/m) \| \to \max_{t_0 \le t \le t_1} \| \mathbf{V}(t) \| \qquad \text{in distribution,} \tag{3.7}$$

where $\mathbf{V}$ is a two-dimensional Gaussian process with mean $\mathbf{0}$ and a covariance matrix,

$$\text{Cov}\,[\mathbf{V}(t), \mathbf{V}(s)] = \begin{pmatrix} I_2 & \Lambda_{ts} \\ \Lambda_{ts} & I_2 \end{pmatrix},$$

I_2 is an identity matrix and

$$\Lambda_{ts} = \begin{pmatrix} \lambda_{11}(t,s) & \lambda_{12}(t,s) \\ \lambda_{21}(t,s) & \lambda_{22}(t,s) \end{pmatrix}$$

with

$$\lambda_{11}(t,s) = \left\{\frac{t(1-s)}{s(1-t)}\right\}^{\frac{1}{2}} \frac{D(s,t)}{\{D(t,t)D(s,s)\}^{\frac{1}{2}}}$$

$$\lambda_{22}(t,s) = \left\{\frac{t^3(1-s)^3}{s^3(1-t)^3}\right\}^{\frac{1}{2}} \frac{D(s,t)}{\{D(t,t)D(s,s)\}^{\frac{1}{2}}}$$

$$\lambda_{12}(t,s) = \left\{3\frac{t(1-s)^3}{s^3(1-t)}\right\}^{\frac{1}{2}} \frac{t-s}{\{D(t,t)D(s,s)\}^{\frac{1}{2}}}$$

$$\lambda_{21}(t,s) = \left\{3\frac{t^3(1-s)}{s(1-t)^3}\right\}^{\frac{1}{2}} \frac{s-t}{\{D(t,t)D(s,s)\}^{\frac{1}{2}}}$$

and

$$D(s,t) = 1 - 3s(1-t) \qquad \text{for } t < s.$$

Proof : In proving this result, we have only to show that

$$\mathbf{V}_m \to \mathbf{V} \qquad \text{in distribution.}$$

To prove that the finite-dimensional distributions of $\mathbf{V}_m$ converge to those of $\mathbf{V}$, it suffices to show that for any sets of $(r_1, \ldots, r_n)$ and $(i_1, \ldots, i_n)$ such that $(i_1/m, \ldots, i_n/m) \to (t_1, \ldots, t_n)$ as $m \to \infty$,

$$E[\exp\{i\sum_{k=1}^{n}(r_{1,k}V_{1,m}(i_k/m) + r_{2,k}V_{2,m}(i_k/m))\}]$$
$$\to E[\exp\{i\sum_{k=1}^{n}(r_{1,k}V_1(t_1) + r_{2,k}V_2(t_k))\}] \qquad \text{in distribution,}$$

which follows from the same argument in Theorem 2.2.6. In Theorem 2.2.6, we have shown that the sequence $\{V_{1,m}\}$ is tight and the similar argument shows that the sequence $\{V_{2,m}\}$ is also tight. Lemma 2.2.5 now implies that the sequence $\{\mathbf{V}_m = (V_{1,m}, V_{2,m})\}$ is tight. And hence $\mathbf{V}_m \to \mathbf{V}$ in distribution. Therefore (3.7) follows from the continuous mapping theorem. ∎

The rest of this section is devoted to a generalization of Theorem 3.2.5 to the case where $x_j = f(j/m)$ for some integrable function f. In fact, we need only to figure out the limiting covariance function to find the limiting distribution of the MLRS, which is described in the following theorem. The proof will be omitted since it is similar as that of Theorem 3.2.5.

Theorem 3.2.6.

Suppose that $x_j = f(j/m)$ $j = 1, \ldots, m$, for some integrable function f such that $f(0) = 0$ and $f(1) = 1$.

Under H_0, as $m \to \infty$ and $m_i/m \to t_i$ for $i = 0, 1$,

$$M_3 = \max_{m_0 \le i \le m_1} \| \mathbf{V}_m(i/m) \| \to \max_{t_0 \le t \le t_1} \| \mathbf{V}(t) \| \qquad \text{in distribution,}$$

where $\mathbf{V}$ is a two-dimensional Gaussian process with mean $\mathbf{0}$ and covariance matrix

$$\mathrm{Cov}\,[\mathbf{V}(t), \mathbf{V}(s)] = \begin{pmatrix} I_2 & \Lambda_{ts} \\ \Lambda_{ts} & I_2 \end{pmatrix},$$

with

$$\lambda_{11}(t,s) = \left\{ \frac{t(1-s)}{s(1-t)} \right\}^{\frac{1}{2}} \frac{D(s,t)}{\{D(t,t)D(s,s)\}^{\frac{1}{2}}}$$

$$\lambda_{22}(t,s) = \left\{ \frac{h(t)[D(s,s) - h(s)]}{h(s)[D(t,t) - h(t)]} \right\}^{\frac{1}{2}} \frac{D(s,t)}{\{D(t,t)D(s,s)\}^{\frac{1}{2}}}$$

$$\lambda_{12}(t,s) = \left\{ \frac{t[D(s,s) - h(s)]}{h(s)(1-t)} \right\}^{\frac{1}{2}} \frac{(1-s)g(s) - (1-t)g(t)}{\{D(t,t)D(s,s)\}^{\frac{1}{2}}}$$

$$\lambda_{21}(t,s) = \left\{ \frac{h(t)(1-s)}{s[D(t,t) - h(t)]} \right\}^{\frac{1}{2}} \frac{sg(s) - tg(t)}{\{D(t,t)D(s,s)\}^{\frac{1}{2}}}$$

and

$$g(t) = \frac{\int_0^1 f(u)du - \left(\int_0^t f(u)du\right)/t}{\left\{(1-t)\left[\int_0^1 f^2(u)du - \left(\int_0^1 f(u)du\right)^2\right]\right\}^{\frac{1}{2}}}$$

$$h(t) = \frac{\int_0^t f^2(u)du - \left(\int_0^t f(u)du\right)^2/t}{\left[\int_0^1 f^2(u)du - \left(\int_0^1 f(u)du\right)^2\right]}$$

$$D(s,t) = 1 - s(1-t)g(s)g(t) \qquad \text{for } t < s.$$

Remark 3.1. In Chapter 2, it was discussed that covariance function depends on the configuration of the values of the independent variable only through the function g. In the case where both the intercept and the slope change, one more function h is involved to explain such a dependence. Also $g(t) = \sqrt{3}$ and $h(t) = t^3$ when $f(u) = u$, which is the case in which $x_j = j/m$.

When the variance is unknown, we obtain the same limiting distribution of M_4 as that of M_3, which is stated in the following corollary.

Corollary 3.2.7.

Suppose that $x_j = f(j/m)\ j = 1, \ldots, m$, for some integrable function f such that $f(0) = 0$ and $f(1) = 1$.

Under H_0, as $m \to \infty$ and $m_i/m \to t_i$ for $i = 0, 1$,

$$M_4 = \max_{m_0 \le i \le m_1} \| \tilde{\mathbf{V}}_m(i/m) \| \to \max_{t_0 \le t \le t_1} \| \mathbf{V}(t) \| \qquad \text{in distribution,}$$

where $\mathbf{V}$ is a two-dimensional Gaussian process defined in Theorem 3.2.6.

3.3. Approximations to Significance Levels

Now our concern is how to approximate significance levels of M_3 and M_4. We follow the basically same arguments used in Section 2.3, extended to boundary crossing problems by a discrete stochastic process which has two-dimensional state space and one-dimensional time parameter. In Section 3.3.1, we give an asymptotic expression which can be used to approximate $\Pr\{M_3 < b\}$ when the x_j's are random, using the argument developed in Siegmund (1986, Chapter 5). Then we derive an approximation to the right-hand tail of the distribution under H_0 of M_3 when the x_j's are fixed. Since these tail probabilities are interpreted as significance levels, it is important that they be accurate when the true probabilities are in the range .01 - .10. We perform Monte carlo experiments and discuss how accurately the asymptotic expressions approximate the actual distribution. Also it will be discussed how well significance levels of M_4 which is the MLRS when σ^2 is unknown, can be approximated by the asymptotic result derived for the known variance case.

3.3.1. When the independent variable is random

In Section 3.2, we showed that

$$M_3 \rightarrow \max_{t_0 \le t \le t_1} \frac{\| \mathbf{W}^0(t) \|}{\{t(1-t)\}^{\frac{1}{2}}} \qquad \text{in distribution,}$$

where $\mathbf{W}^0$ is a two-dimensional Brownian bridge process on [0,1] and $m_i/m \rightarrow t_i$, for $i = 0, 1$, as $m \rightarrow \infty$. In principle, we can approximate the significance level of the test, $\Pr\{M_3 > b\}$, by the tail probability of this limiting distribution. James, James, and Siegmund(1987) give an approximation to

$$\Pr\{\max_{t_0 \le t \le t_1} \| \mathbf{W}^0(t) \| / \{t(1-t)\}^{\frac{1}{2}} > b\},$$

where $\mathbf{W}^0$ is a d-dimensional Brownian bridge process. As in Section 2.3.1, the approximations to tail probabilities of M_3 by those of this limiting distribution are too crude. Since the exact distribution of M_3 is too complicated, we shall now consider an analogous *discrete time result* as in Section 2.3.1. In the following proposition, we derive an approximation to the tail probability defined in terms of a Brownian bridge process observed at discrete instants of time, which is a generalization of (3.12) in Siegmund (1986).

Let $T = \inf\{n : n \ge m_0, \|\mathbf{S}_n\| \ge b\{n(1-n/m)\}^{\frac{1}{2}}\}$, where $\mathbf{S}_n = \mathbf{z}_1 + \cdots + \mathbf{z}_n$ and $\mathbf{z}$'s are independently normally distributed d-dimensional random variables with mean $\mathbf{0}$ and indentity covariance matrix. And let $\Pr_0^{(m)}\{A\} = \Pr\{A \,|\, \mathbf{S}_m = 0\}$.

Proposition 3.3.1.

Assume that $b \rightarrow \infty$, $m_0 \rightarrow \infty$, $m_1 \rightarrow \infty$, $m \rightarrow \infty$ in such a way that for some $0 < t_0 < t_1 < 1$ and $b > 0$,

$$m_i/m \rightarrow t_i \;\text{ for } i = 0, 1, \text{ and } \; b^2/m \rightarrow a.$$

Then as $m \rightarrow \infty$,

$$\Pr_0^{(m)}\{T \le m_1\} \sim b^d e^{-\frac{b^2}{2}} 2^{(1-\frac{d}{2})}[\Gamma(d/2)]^{-1} \int_{b(m_1^{-1}-m^{-1})^{\frac{1}{2}}}^{b(m_0^{-1}-m^{-1})^{\frac{1}{2}}} r^{-1}\nu(r + a/r)dr \qquad (3.8)$$

Proof : (3.8) follows from the extension of the argument in the proof of Theorem 3.11 in Siegmund (1986). ∎

For $d = 2$ we obtain the desired result. Table 7 indicates the accuracy of this asymptotic expression to approximate significance levels of M_3 and M_4 when the independent variable is random. This asymptotic expression gives a crude idea about the significance levels and improves the continuous approximation substantially.

3.3.2. When the independent variable is fixed

When the independent variable is fixed, the MLRS involves a two-dimensional Gaussian process with covariance function given in (3.4). Since the Gaussian process involved is again non-differentiable and non-stationary, we follow the same ideas as in Section 2.3.2 to derive an asymptotic expression for $\Pr\{M_3 > b\}$. However the situation is more complicated than in Section 2.3.2, since we have to deal with two dimensional Gaussian process whose coordinates have non zero covariance.

The following lemma reduces this boundary crossing problem by Gaussian process which has one-dimensional time parameter and two-dimensional state space to the problem involving Gaussian process with one-dimensional time parameter and state space, so that the derivation of an asymptotic expression follows from modifications of the calculations in the one-dimensional case.

Lemma 3.3.2.

Let $\{\mathbf{V}(t) = (V_1(t), V_2(t))\}$ be a two-dimensional stochastic process. Then

$$\begin{aligned}&\Pr\{\max_{m_0 \le i \le m_1} \|\mathbf{V}(i/m)\| \ge b\} \\ &= \Pr\{\max_{m_0 \le i \le m_1} \sup_{0 \le \theta \le 2\pi} [\cos\theta V_1(i/m) + \sin\theta V_2(i/m)] > b\}\end{aligned} \tag{3.9}$$

Proof : Note that

$$\begin{aligned}\cos\theta V_1(i/m) + \sin\theta V_2(i/m) &= \langle(\cos\theta, \sin\theta), (V_1(i/m), V_2(i/m))\rangle \\ &= \|\mathbf{V}(i/m)\| \cos\omega.\end{aligned}$$

where $\langle \cdot, \cdot \rangle$ is an inner product of two vectors and ω is the angle between $(\cos\theta, \sin\theta)$ and $(V_1(i/m), V_2(i/m))$. Then taking the supremum over $0 \le \omega \le 2\pi$, (3.9) holds. ∎

We are now in a position to modify arguments used in Section 2.3.2. To begin, we consider the case where $x_j = j/m$. From Lemma 3.3.3 through Theorem 3.3.7, it is assumed $x_j = j/m$ for $j = 1, \ldots, m$, $Z_m(i/m, \theta) = \cos\theta V_{1,m}(i/m) + \sin\theta V_{2,m}(i/m)$. In Section 2.3.2, $\mathrm{Cov}[U_m(t+h), U_m(t)] = C(t)|h| + o(h)$, so that we took the distance between points of the grid, h, as $1/m$ to make $b^2 h \propto a$. Note, however, that

$$\mathrm{Cov}[Z_m(t+h, \theta+\delta), Z_m(t, \theta)] - 1 = C_1(t,\theta)|h| + C_2(t,\theta)\delta^2 + o(h) + o(\delta^2).$$

Thus under the assumption that $b^2/m \to a$ as $m \to \infty$ and $b \to \infty$, we take h and δ such that $b^2 h \propto a$ and $b\delta \propto a$, so that $h \propto \delta^2$. Hence we use the normalized process $Z_{b,m}^{t,\theta}(i,c) = b(Z_m(t + i/m, \theta + c/\sqrt{m}) - b)$, where $b^2/m \to a$.

Lemma 3.3.3.

Suppose that $x_j = j/m$ for $j = 1, \ldots, m$.

Let $Z_m(i/m, \theta) = \cos\theta V_{1,m}(i/m) + \sin\theta V_{2,m}(i/m)$, and

$Z_{b,m}^{t,\theta}(i,c) = b(Z_m(t + i/m, \theta + c/\sqrt{m}) - b)$, where $b^2/m = a$.

Then as $m \to \infty$ and $b \to \infty$,

$$\begin{aligned} &E\,[\, Z_{b,m}^{t,\theta}(i,c) - x \mid Z_{b,m}^{t,\theta}(0,0) = x \,] = -\mu_a(t,\theta)i - ac^2/2 + o(1), \\ &\mathrm{Cov}\,[\, Z_{b,m}^{t,\theta}(i_1,c_1) - x,\; Z_{b,m}^{t,\theta}(i_2,c_2) - x \mid Z_{b,m}^{t,\theta}(0,0) = x \,] \\ &\qquad\qquad = 2\mu_a(t,\theta)\min(i_1,i_2) + c_1 c_2 a + o(1), \end{aligned}$$

where

$$\mu_a(t,\theta) = \frac{\{[1 - 6t(1-t)]\sin^2\theta - \sqrt{3}(2t-1)\cos\theta\sin\theta + (1/2)\}a}{t(1-t)D(t,t)}, \tag{3.10}$$
$$D(t,t) = 1 - 3t(1-t).$$

Proof: These results follow from straightforward calculations. ∎

Lemma 3.3.4.

Fix n, h, and a. Then for each (t,θ) there is a constant $H_a(t,\theta,n,h) < \infty$ such that, if $b \to \infty$, $m \to \infty$, and $b^2/m \to a$, then

$$\Pr\{\max_{0\le i\le n} \sup_{0\le c\le h} Z_m(t+i/m, \theta + c/\sqrt{m}) \ge b\}/[\frac{\phi(b)}{b}] \to 1 + H_a(t,\theta,n,h),$$

where

$$H_a(t,\theta,n,h) = \int_{-\infty}^{0} \exp(-x)\Pr\{\max_{0\le i\le n} Y_a^{t,\theta}(i) + \sup_{0\le c\le h} S_a(c) \ge -x\}dx,$$

and $Y_a^{t,\theta}(i)$ is a partial sum of *i.i.d.* normal random variables with mean $-\mu_a(t,\theta)$ and variance $2\mu_a(t,\theta)$,

$S_a(c) = c\sqrt{a}S_1 - c^2a/2$ with $S_1 \sim N(0,1)$

and $\{Y_a^{t,\theta}(i)\}$ and $\{S_a(c)\}$ are independent.

Proof : By the previous lemma, the limiting process can be represented as

$$Y_a^{t,\theta}(i) + S_a(c) = [\sigma_a(t,\theta)W(i) - \mu_a(t,\theta)i] + [c\sqrt{a}S_1 - c^2a/2],$$

where W is a standard Brownian motion and $\sigma_a^2(t,\theta) = 2\mu_a(t,\theta)$. Then, following the same argument as in Lemma 12.2.3 of Leadbetter, Lindgren, and Rootzen (1983),

$$\Pr\{\max_{0\le i\le n} \sup_{0\le c\le h} Z_m(t+i/m, \theta + c/\sqrt{m}) > b\}/[\frac{\phi(b)}{b}] \to 1 + H_a(t,\theta,n,h),$$

where c takes real values. ∎

Lemma 3.3.5.

For each (t,θ), there exists a function $H_a^*(t,\theta)$ such that

$$\lim_{\substack{n\to\infty \\ h\to\infty}} H_a(t,\theta,n,h)/(nh) = H_a^*(t,\theta) \qquad \text{uniformly in t and } \theta.$$

As $b \to \infty$ and $m \to \infty$,

$$\Pr\{\max_{t_0\le i/m\le t_1} \sup_{0\le c/\sqrt{m}\le 2\pi} Z_m(i/m, c/\sqrt{m}) \ge b\}/\ [b^2\phi(b)] \to \int_{t_0}^{t_1}\int_0^{2\pi} H_a^*(t,\theta)d\theta dt/a^{\frac{3}{2}}.$$

Proof : Let n and h be fixed integers and

$$B_{k,l} = \{\max_{kn\leq i\leq(k+1)n} \sup_{lh\leq c\leq(l+1)h} Z_m(i/m, c/\sqrt{m}) \geq b\}$$
$$= \{\max_{0\leq i\leq n} \sup_{0\leq c\leq h} Z_m((kn+i)/m, (lh+c)/\sqrt{m}) \geq b\}$$

Then it can be shown that

$$\Pr\{\max_{t_0\leq i/m\leq t_1} \sup_{0\leq c/\sqrt{m}\leq 2\pi} Z_m(i/m, c/\sqrt{m}) \geq b\} \sim \sum_{k=K_0}^{K_1}\sum_{l=L_0}^{L_1} P\{B_{k,l}\},$$

where $K_0 n = m_0$, $K_1 n = m_1$, $L_0 = 0$, $L_1 = 2\pi\sqrt{m}/h$, $K_1 - K_0 = \lfloor m/n \rfloor$, and $L_1 - L_0 = \lfloor 2\pi\sqrt{m}/h \rfloor$.

Now Lemma 3.3.4 implies that

$$\sum_{k=K_0}^{K_1}\sum_{l=L_0}^{L_1} P\{B_{k,l}\} \sim [\phi(b)/b] \sum_{k=K_0}^{K_1}\sum_{l=L_0}^{L_1} [1 + H_a(kn/m, lh/\sqrt{m}, n, h)]$$
$$\sim b^2\phi(b)[2\pi + \sum_{k=K_0}^{K_1}\sum_{l=L_0}^{L_1} H_a(kn/m, lh/\sqrt{m}, n, h)]/(nha\sqrt{a}).$$

Therefore

$$\Pr\{\max_{t_0\leq i/m\leq t_1} \sup_{0\leq c/\sqrt{m}\leq 2\pi} Z_m(i/m, c/\sqrt{m}) > b\}/[b^2\phi(b)]$$
$$\sim \lim_{\substack{n\to\infty\\ h\to\infty}} \int_{t_0}^{t_1}\int_0^{2\pi} H_a(t,\theta,n,h)\,d\theta dt/(nha\sqrt{a})$$
$$= \int_{t_0}^{t_1}\int_0^{2\pi} H_a^*(t,\theta)\,d\theta dt/(a\sqrt{a}),$$

which completes the proof. ∎

Lemma 3.3.6.

For each fixed (t, θ),

$$\int_{t_0}^{t_1}\int_0^{2\pi} H_a^*(t,\theta)\,d\theta dt/a^{\frac{3}{2}} = \int_{t_0}^{t_1}\int_0^{2\pi} \mu_a(t,\theta)\nu[2\mu_a^*(t,\theta)]\,d\theta dt\ /a^{\frac{3}{2}}, \qquad (3.11)$$

where $\mu_a(t,\theta)$ was defined in (3.10) and $\mu_a^*(t,\theta) = \{\mu_a(t,\theta)/2\}^{\frac{1}{2}}$.

Proof : Note that

$$H_a(t,\theta,n,h) = \int_0^\infty \exp(x)\Pr\{\max_{0\leq i\leq n} Y_a^{t,\theta}(i) + \sup_{0\leq c\leq h} S_a(c) \geq x\}dx.$$

where $Y_a^{t,\theta}(i)$ and $S_a(c)$ have the same representation as in Lemma 3.3.4.

Let

$$\Pr\{\max_{0\le i\le n} Y_a^{t,\theta}(i) + \sup_{0\le c\le h} S_a(c) \ge x\} = 1 - R(x).$$

Then

$$\begin{aligned} H_a(t,\theta,n,h) &= \int_0^\infty \exp(x)[1-R(x)]dx \\ &= \int_0^\infty \exp(x)\int_x^\infty dR(y)dx \\ &= \int_0^\infty \int_0^y \exp(x)dxdR(y) \\ &= \int_0^\infty \exp(y)dR(y) - 1 \\ &= \int_0^\infty \exp(y)dF(y)\int_0^\infty \exp(y)g(y)dy - 1 \\ &= \{\int_0^\infty \exp(y)[1-F(y)]dy + 1\}\{\int_0^\infty \exp(y)g(y)dy\} - 1, \end{aligned}$$

where

$$1 - F(y) = \Pr\{\max_{0\le i\le n} Y_a^{t,\theta}(i) \ge y\}$$
$$1 - G(y) = \Pr\{\sup_{0\le c\le h} S_a(c) \ge y\}.$$

By the same argument as in Lemma 2.3.4, as $n \to \infty$,

$$\int_0^\infty \exp(y)[1-F(y)]dy/n \to \mu_a(t,\theta)\nu[2\mu_a^*(t,\theta)].$$

And it can be shown that, as $h \to \infty$,

$$\int_0^\infty \exp(y)g(y)dy/h \to \{a/(2\pi)\}^{\frac{1}{2}},$$

using

$$g(y) = \begin{cases} 1/2, & \text{if } y = 0 \\ \phi(\sqrt{2y})/\sqrt{2y} & \text{if } 0 < y < h^2a/2 \\ \phi(y/(h\sqrt{a}) + h\sqrt{a}/2)/(h\sqrt{a}), & \text{if } y \ge h^2a/2. \end{cases}$$

Then

$$H_a(t,\theta,n,h)/(nha^{\frac{3}{2}}) \to \mu_a(t,\theta)\nu[2\mu_a^*(t,\theta)]/(a\sqrt{2\pi}),$$

as $n, h \to \infty$. ∎

Theorem 3.3.7.

Assume that $b \to \infty$, $m_0 \to \infty$, $m_1 \to \infty$, and $m \to \infty$ in such a way that for some $0 < t_0 < t_1 < 1$ and $a > 0$

$$m_i/m \to t_i, \quad i = 0, 1 \text{ and } b^2/m \to a.$$

Then as $m \to \infty$,

$$Pr\{\max_{m_0 \le i \le m_1} \|\mathbf{V}_m(i/m)\| \ge b\} \sim b^2\phi(b) \int_{t_0}^{t_1} \int_0^{2\pi} \nu[2\mu_a^*(t,\theta)]\mu_a(t,\theta)dt/(a\sqrt{2\pi}), \tag{3.12}$$

where $\mu_a(t,\theta)$ is defined in (3.10) and $\mu_a^*(t,\theta)$ is defined in (3.11).

Table 8 indicates the accuracy of (3.12). From these numerical results, it can be confirmed that (3.12) is quite an accurate approximation to the significance level of M_3 and also gives a reasonable approximation to the tail probability of the null distribution of M_4. In the rest of this section, we generalize Theorem 3.3.7 to the values of the x_j's which satisfy some mild conditions. Proofs will be omitted since they follow closely those of the previous theorem.

Lemma 3.3.8.

Suppose that $x_j = f(j/m)$, $j = 1, \ldots, m$, for some integrable function f such that $f(0) = 0$ and $f(1) = 1$.

Let $Z_m(i/m, \theta) = \cos\theta V_{1,m}(i/m) + \sin\theta V_{2,m}(i/m)$, and

$Z_{b,m}^{t,\theta}(i,c) = b(Z_m(t + i/m, \theta + c/\sqrt{m}) - b)$, where $b^2/m = a$.

Then as $m \to \infty$ and $b \to \infty$,

$$E[Z_{b,m}^{t,\theta}(i,c) - x | Z_{b,m}^{t,\theta}(0,0) = x] = -\mu_a(t,\theta)i - ac^2/2 + o(1),$$

$$\begin{aligned}\mathrm{Cov}[Z_{b,m}^{t,\theta}(i_1,c_1) - x,\ Z_{b,m}^{t,\theta}(i_2,c_2) - x | Z_{b,m}^{t,\theta}(0,0) = x]& \\ = 2\mu_a(t,\theta)\min(i_1,i_2) + c_1c_2a + o(1),&\end{aligned}$$

where

$$\mu_a(t,\theta) = \{1/[t(1-t)] + \sin^2(\theta)A_1(t) - \cos\theta\sin\theta A_2(t)\}a/[2D(t)] \tag{3.13}$$

$$A_1(t) = \frac{h'(t)[D(t)]^2 + 2h(t)g(t)[h(t)g(t) - tD(t)E(t) - [h(t)^2g(t)D(t)]}{h(t)(D(t) - h(t))} - \frac{D(t)}{t(1-t)},$$

$$A_2(t) = 2[h(t)g(t) - tD(t)E(t)]/\{t(1-t)h(t)[D(t) - h(t)]\}^{\frac{1}{2}},$$

$$D(t) = 1 - g^2(t)t(1-t),$$

$$E(t) = g(t) - (1-t)g'(t).$$

Proof : A straight forward calculation suffices. ∎

Theorem 3.3.9.

Suppose that $x_j = f(j/m)$, $j = 1, \ldots, m$, for some integrable function f such that $f(0) = 0$ and $f(1) = 1$.

Assume that $b \to \infty$, $m_0 \to \infty$, $m_1 \to \infty$, *and* $m \to \infty$ in such a way that for some $0 < t_0 < t_1 < 1$ and $a > 0$

$$m_i/m \to t_i, \quad i = 0, 1 \text{ and } b^2/m \to a.$$

Then as $m \to \infty$,

$$Pr\{\max_{m_0 \le i \le m_1} \|\mathbf{V}_m(i/m)\| \ge b\} \sim b^2\phi(b) \int_{t_0}^{t_1} \int_0^{2\pi} \nu[2\mu_a^*(t,\theta)]\mu_a(t,\theta)d\theta dt/(a\sqrt{2\pi}), \tag{3.14}$$

where $\mu_a(t,\theta)$ is defined in (3.13) and $\mu_a^*(t,\theta) = \{\mu_a(t,\theta)/2\}^{\frac{1}{2}}$.

In the case of $x_j = f(j/m)$, $\mu_a(t,\theta)$ involves two different functions h and g through which the distribution of the test statistic depends on the configuration of the x_j's. As a matter of calculation, this case is more complicated than the case of the uniformly spaced x_j's. However previous Monte Carlo experiments lead us to expect (3.14) to be quite good approximations. In this chapter, we have not considered powers and confidence regions. For a confidence region of the change point the method of Cox and Spijøtvoll (1981) can be used, and the argument in Section 2.4 might lead us to a generalization of approximations to powers and confidence regions.

Chapter 4

Concluding Remarks

As discussed in Chapter 1, the exact null distributions of most of the likelihood ratio statistics are too complicated to deal with. Most of previous works have been done by numerical or Monte Carlo methods, e.g. Quandt (1958), Beckman and Cook (1979), Maronna and Yohai (1978), etc. An analytic approach was taken by Worseley (1983) who derived approximations to upper bounds of the null distribution functions of the likelihood ratio statistics.

An important characteristic of the tests considered in Chapters 2 and 3 is that they involve Gaussian processes. Using methods developed to solve boundary crossing problems by a Gaussian process we derived quite accurate approximations to significance levels in various cases. The models that we studied are simple linear regression models. Although we do not consider more complicated models and related problems like confidence regions in general cases, this dissertation may give some insight into those problems. Note that in both (2.16) and (3.12), $b\phi(b)\int_{t_0}^{t_1} \nu[2\mu_a^*(t,\cdot)]\mu_a(t,\cdot)dt/a$ accounts for the boundary crossing probabilities by the given Gaussian processes with respect to time and the integration with respect to the angle θ is involved in (3.12) basically because of the angle parameter introduced to reduce the 2-dimensional problem to the one-dimensional case. This comparison may lead to a generalization of our results. In testing for a change in the coefficient of the multiple regression model, the MLRS is the maximum norm of a d-dimensional Gaussian process. By the same argument in Lemma 3.3.1, we can convert this boundary crossing problem by a d-dimensional Gaussian process to a one-dimensional problem with additional angle parameters. Thus, in general, once the covariance function of the Gaussian process is evaluated, a similar argument may be applied to find asymptotic expressions to approximate significance levels. In our models, the change point is assumed to be one of the data points. Thus our model might be suitable to a set of data which involves discrete time such as annual gross domestic product, number of accidents in consecutive years, and so on. Hinkely (1971) studied a set of data obtained from the experiment to determine the relationship between blood factor VII production and wafarin concentration. In such a case, it is more reasonable to consider a continuous model that a change occurs at some point in the range of the independent variables and two-phase regression line is continuous. Also this example gives a good explanation why we need to think about the two-phase regression model rather than some alternative such as parabolic one.

In many cases a two-phase regression can only be a reasonable approximation, adequate for many purposes. However it is also important to find an appropriate model. As Beckman and Cook (1979) pointed out by example, the continuity assumption may lead to very different estimates of the parameters. The choice of the model is to some extent a matter of experience and common sense. Even though the model should be decided from the biological, economic, or some other particular viewpoint, our model can be applied to give some insight into the decision of a change in the regression relationship and our approximations can be used as convenient standards.

Table 1. Approximations to $\Pr\{M_1 > b\}$

: When Only the Intercept Changes

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

m	b	p_1	p_2	True probability
10	1.9571	0.4936	0.2656	0.25
	2.3854	0.2393	0.1037	0.10
	2.6595	0.1340	0.0511	0.05
	3.1492	0.0384	0.0118	0.01
15	2.0171	0.4522	0.2996	0.25
	2.4412	0.2142	0.1178	0.10
	2.7224	0.1159	0.0568	0.05
	3.2220	0.0312	0.0126	0.01
20	2.0632	0.4215	0.2657	0.25
	2.4733	0.2006	0.1080	0.10
	2.7321	0.1133	0.0556	0.05
	3.2963	0.0250	0.0102	0.01
30	2.1190	0.3856	0.2634	0.25
	2.5253	0.1800	0.1077	0.10
	2.8006	0.0961	0.0529	0.05
	3.3963	0.0185	0.0086	0.01
40	2.1487	0.3672	0.2645	0.25
	2.5598	0.1672	0.1068	0.10
	2.8429	0.0866	0.0518	0.05
	3.3241	0.0230	0.0121	0.01
70	2.2092	0.3313	0.2602	0.25
	2.6274	0.1441	0.1027	0.10
	2.9131	0.0726	0.0487	0.05
	3.4527	0.0155	0.0093	0.01

$m_0 = 0.1 * m, \qquad m_1 = 0.9 * m$

p_1 : Approximations by A.2.3

p_2 : Approximations by A.2.4

Table 2. Approximations to $\Pr\{M_2 > b\}$

: When Only the Intercept Changes

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

m	b	p	(p′)	True probability
10	2.1732	0.1888	(0.1695)	0.25
	2.4692	0.0587	(0.0842)	0.10
	2.6146	0.0256	(0.0577)	0.05
	2.8404	0.0035	(0.0306)	0.01
15	2.1672	0.2475	(0.2203)	0.25
	2.5133	0.0860	(0.0985)	0.10
	2.6955	0.0415	(0.0611)	0.05
	3.0521	0.0058	(0.0217)	0.01
20	2.1703	0.2381	(0.2139)	0.25
	2.5324	0.0883	(0.0935)	0.10
	2.7403	0.0429	(0.0544)	0.05
	3.1280	0.0075	(0.0175)	0.01
30	2.1810	0.2520	(0.2325)	0.25
	2.5769	0.0925	(0.0949)	0.10
	2.8172	0.0435	(0.0506)	0.05
	3.2502	0.0079	(0.0139)	0.01
40	2.2108	0.2484	(0.2335)	0.25
	2.5975	0.0963	(0.0974)	0.10
	2.8312	0.0481	(0.0530)	0.05
	3.2674	0.0100	(0.0145)	0.01
70	2.2431	0.2524	(0.2431)	0.25
	2.6415	0.0988	(0.0991)	0.10
	2.8952	0.0485	(0.0511)	0.05
	3.3743	0.0097	(0.0121)	0.01

$m_0 = 0.1 * m, \qquad m_1 = 0.9 * m$

p : Approximations by A.2.5.

p′ : Approximations by A.2.4.

Table 3. Approximations to $\Pr\{M_i > b_i\}, i = 1,2.$

: When Only the Intercept Changes

$x_j = j/m, \quad j = 1, \ldots, m$

m	b_1	p	b_2	(p′)	True probability
10	2.0756	0.2675	2.2625	(0.1736)	0.25
	2.4519	0.1073	2.4582	(0.0941)	0.10
	2.7157	0.0517	2.6236	(0.0673)	0.05
	3.2747	0.0086	2.8425	(0.0314)	0.01
15	2.1695	0.2769	2.3093	(0.1931)	0.25
	2.5812	0.1005	2.6036	(0.0897)	0.10
	2.8500	0.0468	2.7647	(0.0565)	0.05
	3.3413	0.0095	3.0455	(0.0236)	0.01
20	2.2315	0.2883	2.3395	(0.2136)	0.25
	2.6245	0.1067	2.6553	(0.0946)	0.10
	2.8665	0.0539	2.8388	(0.0559)	0.05
	3.3924	0.0098	3.1921	(0.0183)	0.01
30	2.3271	0.2743	2.3939	(0.2314)	0.25
	2.7133	0.1046	2.7332	(0.0969)	0.10
	2.9598	0.0518	2.9517	(0.0516)	0.05
	3.4829	0.0093	3.3466	(0.0144)	0.01
40	2.3632	0.2846	2.4154	(0.2498)	0.25
	2.7509	0.1083	2.7605	(0.1039)	0.10
	3.0144	0.0510	2.9825	(0.0550)	0.05
	3.5135	0.0099	3.4327	(0.0128)	0.01
70	2.4357	0.2933	2.4635	(0.2736)	0.25
	2.8348	0.1078	2.8328	(0.1076)	0.10
	3.1037	0.0497	3.0712	(0.0543)	0.05
	3.5847	0.0102	3.5166	(0.0128)	0.01

$m_0 = 0.1 * m, \qquad m_1 = 0.9 * m$

b_1 : Percentiles of M_1 (σ^2 is known)

p : Approximations by (2.16)

b_2 : Percentiles of M_2 (σ^2 is unknown)

p′ : Approximations by (2.16)

Table 4. Approximations to $\Pr\{M_1 > b\}$

$: x_j = j/m, \quad j = 1, \ldots, m.$

b\ m	10	20	30	40	50
2.00	0.3150	0.4569	0.5416	0.6002	0.6441
2.05	0.2829	0.4136	0.4922	0.5468	0.5877
2.10	0.2534	0.3733	0.4460	0.4966	0.5347
2.15	0.2263	0.3359	0.4028	0.4496	0.4849
2.20	0.2015	0.3014	0.3628	0.4059	0.4385
2.25	0.1789	0.2696	0.3257	0.3653	0.3953
2.30	0.1584	0.2404	0.2916	0.3278	0.3553
2.35	0.1398	0.2138	0.2603	0.2933	0.3184
2.40	0.1231	0.1896	0.2317	0.2616	0.2845
2.45	0.1081	0.1676	0.2056	0.2327	0.2535
2.50	0.0946	0.1478	0.1819	0.2064	0.2252
2.55	0.0826	0.1300	0.1605	0.1825	0.1995
2.60	0.0719	0.1139	0.1412	0.1610	0.1762
2.65	0.0625	0.0996	0.1239	0.1416	0.1552
2.70	0.0541	0.0868	0.1084	0.1241	0.1363
2.75	0.0468	0.0755	0.0946	0.1085	0.1194
2.80	0.0403	0.0655	0.0823	0.0946	0.1042
2.85	0.0346	0.0566	0.0714	0.0823	0.0908
2.90	0.0297	0.0488	0.0618	0.0714	0.0788
2.95	0.0254	0.0420	0.0533	0.0617	0.0683
3.00	0.0216	0.0360	0.0459	0.0532	0.0590
3.05	0.0184	0.0308	0.0394	0.0458	0.0508
3.10	0.0156	0.0263	0.0337	0.0392	0.0436
3.15	0.0132	0.0223	0.0288	0.0336	0.0374
3.20	0.0111	0.0190	0.0245	0.0286	0.0319
3.25	0.0094	0.0160	0.0208	0.0244	0.0272
3.30	0.0079	0.0135	0.0176	0.0207	0.0231
3.35	0.0066	0.0114	0.0149	0.0175	0.0196
3.40	0.0055	0.0096	0.0125	0.0148	0.0165
3.45	0.0046	0.0080	0.0105	0.0124	0.0139
3.50	0.0038	0.0067	0.0088	0.0104	0.0117
3.55	0.0032	0.0056	0.0073	0.0087	0.0098
3.60	0.0026	0.0046	0.0061	0.0073	0.0082
3.65	0.0026	0.0038	0.0051	0.0061	0.0068
3.70	0.0018	0.0032	0.0042	0.0050	0.0057

Table 4. (Continued)

b\ m	60	70	80	90
2.00	0.6787	0.7069	0.7306	0.7508
2.05	0.6201	0.6465	0.6687	0.6877
2.10	0.5648	0.5895	0.6102	0.6280
2.15	0.5129	0.5358	0.5551	0.5716
2.20	0.4643	0.4856	0.5034	0.5188
2.25	0.4191	0.4387	0.4552	0.4694
2.30	0.3772	0.3952	0.4104	0.4234
2.35	0.3384	0.3550	0.3689	0.3809
2.40	0.3028	0.3179	0.3306	0.3416
2.45	0.2701	0.2838	0.2954	0.3054
2.50	0.2402	0.2527	0.2632	0.2723
2.55	0.2131	0.2243	0.2339	0.2421
2.60	0.1884	0.1986	0.2072	0.2146
2.65	0.1662	0.1753	0.1830	0.1897
2.70	0.1461	0.1543	0.1612	0.1672
2.75	0.1281	0.1354	0.1416	0.1470
2.80	0.1120	0.1185	0.1240	0.1288
2.85	0.0977	0.1034	0.1083	0.1126
2.90	0.0849	0.0900	0.0943	0.0981
2.95	0.0736	0.0781	0.0819	0.0853
3.00	0.0637	0.0676	0.0710	0.0739
3.05	0.0549	0.0584	0.0613	0.0639
3.10	0.0472	0.0502	0.0528	0.0551
3.15	0.0405	0.0431	0.0454	0.0473
3.20	0.0346	0.0369	0.0389	0.0406
3.25	0.0295	0.0315	0.0332	0.0347
3.30	0.0251	0.0268	0.0283	0.0296
3.35	0.0213	0.0228	0.0240	0.0251
3.40	0.0180	0.0193	0.0204	0.0213
3.45	0.0152	0.0163	0.0172	0.0180
3.50	0.0128	0.0137	0.0145	0.0152
3.55	0.0107	0.0115	0.0122	0.0128
3.60	0.0090	0.0097	0.0102	0.0107
3.65	0.0075	0.0081	0.0085	0.0090
3.70	0.0062	0.0067	0.0071	0.0075

$m_0 = 0.1 * m, \qquad m_1 = 0.9 * m$

Table 5. Approximations to $\Pr\{M_i > b_i\},\ i = 1,2$

: When Only the Intercept Changes

$x_j = (j/m)^2,\ \ j = 1,\ldots,m$

m	b_1	p_1	p_2	b_2	(p')	True prob.
10	2.0505	0.2780	0.2740	2.2472	(0.1771)	0.25
	2.4506	0.1061	0.1050	2.4949	(0.1050)	0.10
	2.7066	0.0522	0.0518	2.6179	(0.0518)	0.05
	3.2249	0.0101	0.0100	2.8391	(0.0352)	0.01
20	2.2109	0.2902	0.2886	2.3253	(0.2234)	0.25
	2.6230	0.1055	0.1050	2.6493	(0.0982)	0.10
	2.8938	0.0489	0.0487	2.8336	(0.0584)	0.05
	3.3611	0.0108	0.0108	3.1754	(0.0202)	0.01
30	2.3073	0.2835	0.2827	2.3709	(0.2450)	0.25
	2.7085	0.1045	0.1042	2.7229	(0.1005)	0.10
	2.9652	0.0502	0.0501	2.9413	(0.0539)	0.05
	3.5121	0.0083	0.0083	3.3769	(0.0133)	0.01
40	2.3487	0.2913	0.2908	2.4024	(0.2575)	0.25
	2.7526	0.1065	0.1063	2.7729	(0.1007)	0.10
	3.0146	0.0503	0.0502	2.9980	(0.0528)	0.05
	3.4915	0.0106	0.0106	3.4377	(0.0128)	0.01
50	2.3754	0.2983	0.2979	2.4220	(0.2681)	0.25
	2.7630	0.1140	0.1139	2.7788	(0.1092)	0.10
	3.0010	0.0581	0.0581	3.0078	(0.0569)	0.05
	3.5039	0.0114	0.0114	3.4518	(0.0137)	0.01

$m_0 = 0.1 * m, \qquad m_1 = 0.9 * m$

b_1 : Percentiles of M_1 (σ^2 is known)

p_1, p_2 : Approximations by (2.19)

b_2 : Percentiles of M_2 (σ^2 is unknown)

p' : Approximations by (2.19)

Table 6. Approximations to Powers

: When Only the Intercept Changes

$x_j = j/m, \quad j = 1, \ldots, m$

$\alpha_2 - \alpha_2$	ρ	m	LRT	True prob.	MLRT	True prob.
3.6	10	20	0.8903	0.8854	0.9012	0.8974
	5		0.9794	0.9681	0.9863	0.9720
	3		0.9675	0.9573	0.9746	0.9622
3.0	10	20	0.7157	0.7180	0.7351	0.7361
	5		0.8580	0.8643	0.8708	0.8753
	3		0.8391	0.8428	0.8530	0.8557
2.4	10	20	0.4708	0.4840	0.4942	0.5058
	5		0.6290	0.6568	0.6510	0.6756
	3		0.6045	0.6208	0.6270	0.6420
		Critical value	2.9204	(5%)	2.8665	(5%)
2.4	20	40	0.8800	0.8869	0.8960	0.8988
	10		0.9664	0.9619	0.9756	0.9664
	5		0.9360	0.9374	0.9472	0.9454
1.6	20	40	0.4893	0.5158	0.5230	0.5364
	10		0.6381	0.6757	0.6693	0.7010
	5		0.5768	0.6136	0.6096	0.6486
1.0	20	40	0.1887	0.2494	0.2131	0.2639
	10		0.2608	0.3312	0.2898	0.3516
	5		0.2289	0.2950	0.2561	0.3041
		Critical value	2.8253	(10%)	2.7509	(10%)

Table 7. Approximations to $\Pr\{M_i > b_i\}$, $i = 3,4$

: When Both the Intercept and Slope Change

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

m	b_3	p	b_4	(p′)	True probability
10	2.2803	0.1859	2.4911	(0.1289)	0.25
	2.6949	0.0856	2.7060	(0.0836)	0.10
	2.9598	0.0464	2.8115	(0.0661)	0.05
	3.5379	0.0091	2.9682	(0.0455)	0.01
20	2.5024	0.2101	2.6090	(0.1710)	0.25
	2.9039	0.8976	2.9081	(0.0889)	0.10
	3.1656	0.0463	3.0908	(0.0564)	0.05
	3.6775	0.0100	3.4050	(0.0235)	0.01
30	2.5547	0.2052	2.6373	(0.1745)	0.25
	2.9692	0.0835	2.9784	(0.0817)	0.10
	3.2401	0.0414	3.1834	(0.0483)	0.05
	3.6940	0.0105	3.5136	(0.0187)	0.01
40	2.5794	0.2050	2.6405	(0.1819)	0.25
	2.9644	0.0891	2.9880	(0.0841)	0.10
	3.2474	0.0430	3.2198	(0.0463)	0.05
	3.7426	0.0096	3.6063	(0.0149)	0.01
50	2.6238	0.1943	2.6723	(0.1763)	0.25
	3.0193	0.0808	3.0338	(0.0780)	0.10
	3.2783	0.0410	3.2641	(0.0438)	0.05
	3.7726	0.0090	3.6846	(0.0121)	0.01

$m_0 = 2$, $m_1 = 8$ for $m = 10$

$m_0 = 0.1 * m$, $m_1 = 0.9 * m$ for $m \geq 20$

b_3 : Percentiles of M_3 (σ^2 is known)

p : Approximations by (3.8)

b_4 : Percentiles of M_4 (σ^2 is unknown)

p′ : Approximations by (3.8)

Table 8. Approximations to $\Pr\{M_i > b_i\},\ i = 3, 4$

: When Both the Intercept and Slope Change

$x_j = j/m,\ \ j = 1, \ldots, m.$

m	b_3	p	b_4	(p′)	True probability
10	2.3440	0.2917	2.5149	(0.2186)	0.25
	2.7557	0.1078	2.7157	(0.1352)	0.10
	3.0350	0.0488	2.8207	(0.1032)	0.05
	3.5315	0.0095	2.9741	(0.0679)	0.01
20	2.5893	0.2746	2.6871	(0.2287)	0.25
	2.9803	0.1015	2.9628	(0.1134)	0.10
	3.2348	0.0481	3.1387	(0.0691)	0.05
	3.7381	0.0088	3.4394	(0.0273)	0.01
30	2.6711	0.2877	2.7411	(0.2510)	0.25
	3.0584	0.1065	3.0619	(0.1098)	0.10
	3.3040	0.0516	3.2483	(0.0642)	0.05
	3.7634	0.0110	3.6044	(0.0206)	0.01
40	2.7121	0.3025	2.7772	(0.2649)	0.25
	3.1058	0.1100	3.1156	(0.1101)	0.10
	3.3700	0.0502	3.3256	(0.0596)	0.05
	3.8502	0.0098	3.7341	(0.0156)	0.01
50	2.7636	0.2971	2.8077	(0.2717)	0.25
	3.1519	0.1122	3.1496	(0.1115)	0.10
	3.3863	0.0504	3.3647	(0.0593)	0.05
	3.8851	0.0099	3.7778	(0.0152)	0.01

$m_0 = 2,\qquad m_1 = 8\qquad$ for $m = 10$

$m_0 = 0.1 * m,\qquad m_1 = 0.9 * m\qquad$ for $m \geq 20$

b_3 : Percentiles of M_3 (σ^2 is known)

p : Approximations by (3.12)

b_4 : Percentiles of M_4 (σ^2 is unknown)

p′ : Approximations by (3.12)

Appendices

A.1. Basic Facts about Convergence of Probability Measures

Convergence in distribution of a sequence $\{X_n\}$ of real random variables is traditionally defined to mean convergence of distribution functions at each continuity point of the limit distribution function. For random elements of more general spaces not equipped with a partial ordering, even the concept of distribution function disappears. In Chapter 1 of Billingsley (1968), convergence in distribution for a sequence of random elements was summarized and now we define the convergence in distribution for random elements using his results.

Let $C = C[0,1]$ be the space of continuous functions on [0,1], where we give C the uniform topology by defining the distance between the points x, y as

$$d(x,y) = \sup_{0 \leq t \leq 1} |x(t) - y(t)|.$$

Chapter 2 of Billingsley (1968) contains a theory about the weak convergence in the space C which is used in this dissertation. Here, we include a brief review of definitions and theorems which are basic and important.

Suppose now that $\{X_n\}$ is a sequence of random elements in C. That is, for each ω in Ω, $X_n(\omega)$ is an element of C whose values at t we denote by $X_n(t,\omega)$. For points $t_1, \ldots, t_k$ in $[0,1]$, let $\pi_{t_1,\ldots,t_k}$ be the mapping that carries the point x of C to the point $(x(t_1), \ldots, x(t_k))$ of R^k. The finite dimensional sets are now defined as sets of the form $\pi^{-1}_{t_1,\ldots,t_k} H$ with $H \in R^k$ and the finite dimensional distribution of X_n as that of $\pi_{t_1,\ldots,t_k} X$. Since the space of Borel sets of C with the uniform metric is separable and complete, the finite dimensional sets generate the space of Borel sets. However, the convergence in distribution of $\pi_{t_1,\ldots,t_k} X$ does not imply the convergence of X_n in distribution. The difficulty and interest of weak convergence in C al! come from the fact that it involves considerations going beyond those of finite dimensional sets. Here is an idea which provides a powerful technique for proving weak convergence in C. If every sequence of X_n contains a subsequence which converges in distribution, then X_n converges in distribution. In the space C this condition is equivalent to "tightness" which is a condition that has the effect of preventing the escape of mass to infinity in a certain sense. Now we define tightness of a sequence of random elements as follows: X_n is tight if $X_n(0)$ is tight on line and if for each positive ϵ and η there exists a δ, $o < \delta < 1$, and an integer n_0 such that

$$\frac{1}{\delta} \Pr\{ \sup_{t \leq s \leq t+\delta} |X_n(s) - X_n(t)| \geq \epsilon \} \leq \eta$$

for $n \geq n_0$ and $0 \leq t \leq 1$. Then we have the following result.

Theorem A.1.1

Let $X, X_1, X_2, \ldots,$ be random elements of C. *If the finite dimensional distributions of* X_n converges to those of X, and if $\{X_n\}$ is tight, then $X_n \Rightarrow X$.

To obtain the limiting distributions of the test statistics defined in Sections 2.1 and 3.1 when the independent variables are random, Donsker's theorem was used as an important tool. Donsker formulated a refinement of the central limit theorem by proving weak convergence of the distributions of certain random functions constructed *from the partial* sum.

Theorem A.1.2 (Donsker)

Let $y_1, y_2, \ldots$ be *i.i.d.* random variables with mean 0 and finite, positive variance σ^2, and let $S_n = y_1 + \ldots + y_n$. Define a random element X_n of C by

$$X_n(t, \omega) = \frac{1}{(\sigma\sqrt{n})} S_{[nt]}(\omega) + (nt - [nt]) \frac{1}{(\sigma\sqrt{n})} y_{[nt]+1}(\omega).$$

Then as $n \to \infty$, X_n converges to a Brownian motion process in distribution.

A.2. Applications of Boundary Crossing Probabilities to Change-Point Problems

Methods developed to approximate boundary crossing probabilities in fixed sample statistical problems provide an important tool in this dissertation. Especially, the results in Siegmund (1986) and James, James, and Siegmund (1987) developed for change-point problems were used to approximate the significance levels of the modified likelihood ratio statistics defined in Sections 2.1. and 3.1. The above two papers are concerned with the problem of testing a sequence of normal random variables with constant, known or unknown, variance for no change in mean versus alternatives with a single change-point.

Let $x_1, \ldots, x_m$ be independent random variables and consider the case where the x_n's are normally distributed with mean $\mu^{(n)}$ and constant variance. One specific problem is to test

$$H_0 \ : \mu^{(1)} = \cdots = \mu^{(m)}, \qquad \text{against}$$

$$H_1 \ : \exists 1 \leq \rho < m \qquad \text{such that}$$

$$\mu^{(1)} = \cdots = \mu^{(\rho)} \neq \mu^{(\rho+1)} = \cdots = \mu^{(m)}.$$

When the variance is known, Siegmund (1986) suggests

$$\max_{m_0 \leq k \leq m_1} |S_k - kS_m/m|/\{k(1-k/m)\}^{\frac{1}{2}} \tag{A.2.1}$$

as a test statistic and derives an approximation to the significance level of the test based on (A.2.1). As an application of the theories of weak convergence of stochastic processes ,

$$\Pr\{\max_{m_0 \leq k \leq m_1} |S_k - kS_m/m|/\{k(1-k/m)\}^{\frac{1}{2}} \geq b\} \tag{A.2.2}$$

can be approximated by the corresponding probability defined in terms of a Brownian motion process $W(t)$ $(0 \leq t < \infty)$. That is, (A.2.2) is approximately

$$\begin{aligned} &\Pr\{|W_0(t)| \geq b\{t(1-t)\}^{\frac{1}{2}} \text{ for some } \varepsilon_1 \leq t \leq 1-\varepsilon_2\} \\ &\quad = (b-b^{-1})\phi(b)\log[(1-\varepsilon_1)(1-\varepsilon_2)/\varepsilon_1\varepsilon_2] + 4b^{-1}\phi(b) + o(b^{-1}\phi(b)), \end{aligned} \tag{A.2.3}$$

which is given in Siegmund (1986). The following theorem given in Siegmund (1986) provides an approximation to the significance level of the test statistic (A.2.1) , taking discreteness into consideration.

Theorem A.2.1.

Assume that $b \to \infty$, $m_0 \to \infty$, $m \to \infty$ in such a way that for some $0 \leq t_0 < t_1 < 1$ and $b_0 > 0$

$$m_i/m \to t_i, \qquad i = 0, 1 \quad \text{and} \quad b/\sqrt{m} = b_0.$$

Let $\xi = m\xi_0$ for some $|\xi_0| \in \left(b_0(1-t_1)\{t_0(1-t_0)\}^{\frac{1}{2}}, b_0\{t_1(1-t_1)\}^{\frac{1}{2}}\right)$.

Then as $m \to \infty$,

$$\Pr\{\max_{m_0 \le k \le m_1} |S_k - kS_m/m|/\{k(1-k/m)\}^{\frac{1}{2}} \ge b\}$$
$$\cong 2b\phi(b) \int_{b(m_1^{-1}-m^{-1})^{\frac{1}{2}}}^{b(m_0^{-1}-m^{-1})^{\frac{1}{2}}} x^{-1}\nu(x + b^2/mx)dx + 2[1 - \Phi(b)], \qquad (A.2.4)$$

where ν is given by (2.10).

In the case of unknown and constant variance, James, James, and Siegmund (1987) considered the statistic,

$$\max_{m_0 \le k \le m_1} \left[\frac{|S_k - kS_m/m|}{\{k(1-k/m)\}^{\frac{1}{2}}} \{m^{-1} \sum_{n=1}^{m} (x_n - \bar{x}_m)^2\}^{-\frac{1}{2}} \right]$$

and provides the following approximation which can be used to approximate the significance level.

Corollary A.2.2

Under the same assumptions as in Theorem A.2.1,

$$\Pr\{\max_{m_0 \le k \le m_1} \left[\frac{|S_k - kS_m/m|}{\{k(1-k/m)\}^{\frac{1}{2}}} \{m^{-1} \sum_{n=1}^{m} (x_n - \bar{x}_m)^2\}^{-\frac{1}{2}} \right] \ge b\}$$
$$\cong 2\{m/(2\pi)\}^{\frac{1}{2}} \int_{b_0}^{1} (1-x^2)^{(m-4)/2} dx \qquad (A.2.5)$$
$$+ (2/\pi)^{\frac{1}{2}} b(1 - b^2/m)^{(m-4)/2} \int x^{-1}\nu[x + b^2/\{m(1-b_0^2)x\}]dx,$$

where the second integral on the right side is over

$$(b\{(m_1^{-1} - m^{-1})/(1-b_0^2)\}^{\frac{1}{2}}, b\{(m_0^{-1} - m^{-1})/(1-b_0^2)\}^{\frac{1}{2}}).$$

References

Beals, R. E. (1972) Statistics for Economists. *Rand McNally & Company, Chicago.*

Beckman, R. J., and Cook, R. D. (1979) Testing for two-phase regressions. *Technometrics*, **21**, 65-69.

Billingsley, P. (1968) Convergence of Probability Measures. *John Wiley & Sons, New York.*

Brown, R. L., Durbin, J. and Evans, J. M. (1975) Techniques for testing the constancy of regression relations over time. *J. Roy. Statist. Soc. Ser. B*, **37**, 149-192.

Cox, D. R. and Spijøtvoll, E. (1982) On partitioning means into groups. *Scand. J. Statist.*, **9**, 147-52.

Durbin, J. (1985) The first-passage density of a continuous Gaussian Process to a general boundary *J. Appl. Probab.*, **22**, 99-122.

Esterby, S. R. and Elshaarawi, A. H. (1981) Inference about the point of changes in a regression model. *Appl. Statist.*, **30**, 277-285.

Feder, P. L. (1975a) On asymptotic distribution theory in segmented regression problems-identified case. *Ann. Statist.*, **3**, 49-83.

Feder, P. L. (1975b) The log likelihood ratio in segmented regression. *Ann. Statist.*, **3**, 84-97.

Gardner, L. A. Jr., (1969) On detecting changes in the mean of normal variates. *Ann. Mathe. Statist.*, **40**, 116-126.

Haddad, G. G., Jeng, H., and Lai, T. L. (1987), Heart rate variability during respiratory pauses in puppies and dogs. *Pediatric Research*, **22**, 306-311.

Hawkins, D. M. (1977) Testing a seouence of observations for a shift in location. *J. Amer. Statist. Assn.*, **70**, 180-186.

Hawkins, D. M. (1980) A note on continuous and discontinuous segmented regressions. *Technometrics*, **22**, 443-444.

Hinkley, D. V. (1969) Inference about the intersection in two-phase regression. *Biometrika*, **56**, 495-504.

Hinkley, D. V. (1970) Inference about the change-point in a sequence of random variables. *Biometrika*, **57**, 1-17.

Hinkley, D. V. and Hinkely, E. A. (1970) Inference about the change-point in a sequence of binomial random variables. *Biometrika*, **57**, 477-488.

Hinkley, D. V. (1971) Inference in two-phase regression. *J. Amer. Statist. Assn.*, **66**, 736-743.

Hogan, M. L. and Siegmund, D. (1986) Large deviation for the maxima of some random fields. *Adv. Appl. Math.*, **7**, 2-22.

James, B., James, K. L., and Siegmund, D. (1987) Tests for a change-point, *Biometrika*, **74**, 71-83.

Leadbetter, M. R., Lindgren, G., and Rootzen, H. (1983) Extremes and Related Properties of random Sequences and Processes. *Springer-Verlag, New York.*

Lorden, G. (1971) Procedures for reacting to a change in distribition. *Ann. Math. Statist.*, **42**, 1897-1908.

Matthews, D. E., Farewell, V. T. and Pyke, R. (1985) Asymptotic score-statistic processes and tests for constant hazard against a change-point alternative. *Ann. Statist.* **13**, 583-591.

Maronna, R. and Yohai, V. J. (1978) A bivariate test for the detection of a systematic change in mean. *J. Amer. Statist. Assn.*, **73**, 640-645.

Page, E. S. (1954) Continuous inspection schemes. *Biometrika*, **41**, 100-115.

Pettitt, A. N. (1971) A non-parametric approach to the change-point problem. *Appl. Statist.*, **28**, 126-155.

Pettitt, A. N. (1980) A simple cumulative sum type statistic for the change-point problem with zero-one observations. *Biometrika*, **67**, 79-84.

Picard, D. (1985) Testing and estimating change-points in time series. *Adv. Appl. Probab.*, **17**, 841-867.

Pollak, M. (1985) Optimal detection of a change in dustribution. *Ann. Statist.*, **13**, 206-227.

Quandt, R. E. (1958) The estimation of the parameters of a linear regression system obeying two seperate regimes. *J. Amer. Statist. Assn.*, **53**, 873-880.

Quandt, R. E. (1960) New approach to estimating switching regressions. *J. Amer. Statist. Assn.*, **67**, 306-330.

Shiryayev, A. N. (1963) On optimal methods in earliest detection problems. *Theory Probab. Appl.*, **8**, 26-51.

Siegmund, D. (1982) Large deviations for boundary crossing probabilities. *Ann. Probab.*, **10**, 581-588.

Siegmund, D. (1985) Sequential Analysis : Test and Confidence Intervals. *Springer-Verlag, New York.*

Siegmund, D. (1986) Boundary crossing probabilities and statistical applications. *Ann. Statist.*, **14**, 361-404.

Worsley, K. J. (1983) Testing for a two-phase multiple regression. *Technometrics*, **25**, 35-42.

Worsley, K. J. (1983) The power of likelihood ratio and cumulative sum tests for a change in a binomial probability. *Biometrika*, **70**, 455-64.

Worsley, K. J. (1986) Confidence regions and tests for a change-point ina sequence of exponential family random variables. *Biometrika*, **73**, 91-104.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 7	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (and Subtitle) CHANGE-POINT PROBLEMS IN REGRESSION		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Hyune-Ju Kim		8. CONTRACT OR GRANT NUMBER(s) N00014-87-K-0078
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics - Sequoia Hall, Stanford University, Stanford, CA 94305-4065		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-373
11. CONTROLLING OFFICE NAME AND ADDRESS Statistics & Probability Program Code (411(SP)), Office of Naval Research, Arlington, Virginia 22217		12. REPORT DATE June 1988
		13. NUMBER OF PAGES 80pp
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See reverse side for abstract.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT.

This dissertation focuses on the problem of testing for a change in the regression model when errors are independently, normally distributed with constant, known or unknown variance. First we consider the regression model in which only the intercept changes at some unknown point (Model-1). Secondly, the model in which both intercept and slope change is considered (Model-2). In all cases, the likelihood ratio statistic (LRS) is of the form $U = \max_{1 \leq i < m} U_i$, where distributions of U_i's vary according to the assumptions.

In both models, we consider the likelihood ratio test (LRT) as the problem of the boundary crossing by the discrete stochastic process and study problems such as approximations to significance levels, powers, and confidence regions for a change point. First of all, we propose a modified LRT and discuss asymptotic properties of test statistics in cases of random and fixed independent variables. In both cases, we derive analytical approximations to significance levels. When the independent variables are random, the limiting distribution of the modified LRS is a function of a Brownian motion and approximations in Siegmund (*1986, Annals of Statistics*) are used. For fixed independent variables, the limiting distribution involves a Gaussian process with nondifferentiable sample paths. In this case, an approximation is derived assuming the known variance and mild conditions about the empirical distribution of the independent variable, using the argument in Leadbetter, Lindgren and Rootzen (1983,Chapter12), modified for discrete time by Hogan and Siegmund (1986, Advances in Applied Mathematics). In Model-1, we are also concerned with the power of the LRT and confidence regions for a change point.

Numerical approximations of significance levels and powers of the LRT and the results of corresponding Monte Carlo experiments are obtained. We find that the simulations confirm that the theoretical results perform well and demonstrate that the results also can be applied to the unknown variance case.